

An ailing agency

Public health needs strong advocacy within government — and Congress should make sure that the US Centers for Disease Control and Prevention continues to provide it.

"Right now, we are hard at work doing what we do best — protecting people's health whenever, wherever and however we are needed. For that I, and people around the world, are most grateful." This is what Julie Gerberding, the director of the US Centers for Disease Control and Prevention (CDC), wrote in an e-mail message to employees last week; it was subsequently reprinted in the *Atlanta Journal-Constitution*.

Many of us do indeed have cause to be grateful for the work of the CDC. The federal agency, which started in 1946 as a small office to investigate malaria, now has around 9,000 staff and is dedicated to improving all aspects of public health in the United States. When members of the public are concerned about what vaccinations to give their children, they turn to the CDC for advice. When local health departments need guidance on the spread of HIV, they too look to the Atlanta-based agency.

Outside the United States, the CDC enjoys a hard-won reputation for its knowledge of infectious disease. Take, for example, its unparalleled 121 Cities programme for monitoring influenza, as part of which epidemiologists collect weekly figures on the number of influenza deaths from (as it happens) 122 US metropolitan areas. The programme, which can highlight a particularly pernicious flu season at its outset, is unmatched elsewhere in the world.

But some of the people who respect and rely on the CDC are now expressing worries about its own state of health. Some of those concerns are being expressed at the very roots of the agency itself, by the dedicated public-health workers on whose reputation it was built. It was the *Atlanta Journal-Constitution's* reporting of these concerns that prompted Gerberding's e-mail retort.

The complainants allege that good science at the agency is being hampered by bureaucracy and mismanagement. The problems have arisen in part as the result of a reorganization instigated by Gerberding in 2003, and some officials contend that they are being exacerbated by the Bush administration's efforts to exert political influence over the CDC. Some very senior people are leaving; others say they

are staying only until they can collect a pension (see page 250).

Some of these complaints may well have been provoked by any kind of organizational revamp. But when five former directors of the CDC feel compelled to intervene, as they did in a letter sent to Gerberding last year, it is time for outsiders to pay attention.

The CDC's role in helping to assure public health has never been more important. Emerging infectious diseases such as SARS and avian influenza demand a rapid response, and epidemics of HIV and tuberculosis show few signs of abating. Obesity, heart disease and cancer end too many lives prematurely and demand authoritative and assertive management.

The CDC is not the only globally significant public-health organization whose performance is currently under scrutiny. Later this year, the World Health Organization (WHO) is due to elect a director-general to succeed Lee Jong-wook, who died this summer. It is critically important that the WHO chooses a leader with the political and administrative skills needed to make the organization an even more effective player in addressing global public-health issues. Unfortunately, given the intrigue that often surrounds such contests, close observers of the WHO have scant grounds for optimism that this election will yield such a leader.

The Senate Finance Committee is already looking into alleged staff morale problems at the CDC, as well as the agency's use of funds that it has been asked to spend to counter bioterrorism. It is the duty of congressional committees to ensure that the agency and its money are being competently managed. Scientific bodies such as the National Academies could also be asked to play a role in monitoring the CDC's well-being. They should welcome any opportunity to do so, to help ensure that the agency maintains its proud tradition as an effective champion of public health, at home and abroad. ■

"When five former directors of the CDC feel compelled to intervene, as they did last year, it is time for outsiders to pay attention."

Libya's travesty

Six medical workers in Libya face execution. It is not too late for scientists to speak up on their behalf.

Imagine that five American nurses and a British doctor have been detained and tortured in a Libyan prison since 1999, and that a Libyan prosecutor called at the end of August for their execution by firing squad on trumped-up charges of deliberately contaminating more than 400 children with HIV in 1998. Meanwhile, the international community and its leaders sit by, spectators of a farce of a

trial, leaving a handful of dedicated volunteer humanitarian lawyers and scientists to try to secure their release.

Implausible? That scenario, with the medics enduring prison conditions reminiscent of the film *Midnight Express*, is currently playing out in a Tripoli court, except that the nationalities of the medics are different. The nurses are from Bulgaria and the doctor is Palestinian (see page 254).

Despite the medics' plight, the United States agreed in May to re-establish diplomatic relations with Libya, 18 years after the bombing of an airliner over Lockerbie in Scotland that killed 270 civilians. Many observers had expected a resolution of the medics' case to be part of the deal. And the European Union has given Muammar



Gaddafi, the Libyan leader, red-carpet treatment at the European Commission in Brussels.

International diplomacy, dealing as it does with geopolitical and economic *realpolitik*, by necessity often involves turning a blind eye. But its lack of progress in response to the medics' case in Libya is an affront to the basic democratic principles that the United States and the European Union espouse. Diplomacy has lamentably failed to deliver.

The principles of law and science have the common aim of discovering the truth. A previous assessment of the case by two prominent AIDS researchers, Luc Montagnier and Vittorio Colizzi, concluded that the charges are false, that the medics are innocent, and that the infections resulted from poor hygiene in Libya's hospitals. It was not a plot orchestrated by the CIA and Israel's Mossad, as President Gaddafi alleged in 2001 — an allegation that has driven a popular thirst for vengeance in Libya.

The case is politically embarrassing for Gaddafi. Finding a scapegoat is easier than having to admit that the infection of the children was an accidental tragedy. But the most likely diplomatic compromise — that the medics will be condemned to death, with this being commuted to a life sentence — is unacceptable. They are innocent, and the law and science can prove it, if they get the belated opportunity.

That is why scientists should lend their full support to the call by Lawyers without Borders — a volunteer organization that last year helped win the freedom of Amina Lawal, who had been sentenced to death in Nigeria for having a child outside marriage — that Libya's courts should order a fully independent, international scientific assessment of how the children were contaminated.

In 2004, an Editorial in this journal stated, with respect to the medics' case, that "Gaddafi has a chance to show the world that he now understands that true leadership means embracing justice, compassion and a respect for scientific evidence" (*Nature* **430**, 277; 2004). Two years on, we are still waiting, and Lawyers without Borders is right to hold President Gaddafi and the international community to account.

The scientific community has also been relatively silent on the case, perhaps in the hope that it would be sorted out by diplomacy. But the latter has not proved to be the case, and scientific leaders need to use all their influence urgently, as the fate of the medics will be sealed in the coming weeks. It is time not only to save the doctor and nurses, but also to defend a common vision of science and law in establishing the truth, above all other imperatives. Meanwhile, Gaddafi has the opportunity to put this affair behind him by giving the six an immediate pardon. ■

The brief goodbye

Because of trends in submissions, *Nature's* Brief Communications will bow out at the end of the year.

From time to time these pages have proudly announced the birth of new sections of *Nature* and new research journals. Rarely, if ever, have we proclaimed a death. It is with mixed feelings that we now announce the demise of Brief Communications, whose final appearance will be in the last issue of 2006. Importantly, we will continue to give full support to Brief Communications Arising, an online-only section in which we publish critical discussions of *Nature* papers.

The Brief Communications section has provided a bridge between the journalism and opinionated sections of *Nature* and the review articles and full-scale research papers. The sections' intentions have been to capture the excitement and appeal of both — and in doing so has ensured plenty of impact both among readers and in the media.

There have been tales of rat robots, Neolithic noodles, nano-bulls, a cloned cat (as well as a more notorious cloned dog), of goings-on at the Moon's north pole and with the Queen's vowels, and there have even been models to unravel the best way of tying shoelaces, to list just a few. In generally no more than a single page, papers in the section have offered glimpses into breaking scientific news — including the giant Indonesian earthquake, SARS and bird flu, and on surprising climate effects of shutting down air traffic for three days after the attacks of 11 September 2001. All were underpinned by rigorous peer review, and the success of Brief Communications owes much to our referees entering into the often quirky spirit of the section without compromising *Nature's* standards.

The section has had its critics. Sober scientists have worried that such brief candles have diminished the stellar luminescence of the *Nature* references in their CVs. False rumours that the section was not peer reviewed have occasionally circulated. But *Nature* has stood by its short masterpieces, quirkiness and all.

Why, then, abandon this popular part of the journal, which continues to receive many more submissions than it can possibly publish? It is, we believe, an appropriate response to what we are increasingly receiving, and a belief that the pages can therefore be better deployed. Fewer and fewer submissions to Brief Communications have been making the grade — perhaps constrained by the short format and limited online supplementary material, they may be too lightweight, too technical, too long or too preliminary.

Perhaps today's pressures are forcing science to become more earnest and more specialized, as well as demanding greater detail in presentation. Never mind that Watson and Crick's paper on the structure of DNA was the length of a Brief Communication.

The section's demise will not affect the Brief Communications appearing in *Nature* research journals, where there is undoubtedly a place for short accounts for a more specialized audience. Moreover, *Nature* will, if anything, increase its support for Brief Communications Arising, in which critics of our papers can vent their views, usually with authors' responses.

We also remain committed to another virtue embodied in Brief Communications: being brief. The quality of papers can supposedly be enhanced by boundless web space; length is no longer such an issue. But our authors should be warned, and our readers reassured, that *Nature* remains a champion of succinctness. ■

"Perhaps today's pressures are forcing science to become more earnest and more specialized."

RESEARCH HIGHLIGHTS

ANIMAL BEHAVIOUR

Slow food

J. Mammal. **87**, 790–798 (2006)

The sluggish lifestyle of the slow loris (*Nycticebus coucang*), a small primate weighing about two kilograms, is the product of a metabolism slow enough to rival that of the sloth. Now Frank Wiens of Bayreuth University in Germany and his colleagues have explained how the animal's diet causes its lack of energy.

Wien's team studied the feeding behaviour of lorises in West Malaysia and found that the animals actually eat plenty of sugary food all year round. The problem seems to be that the lorises also eat lots of toxic plant material. Glucose is needed to detoxify these substances, reducing the amount available for the animal's metabolism.



A. SHAH/NATUREPL.COM

BIOCHEMISTRY

Cell detectives

Anal. Chem. doi:10.1021/ac0607010 (2006)

A team from the University of Illinois in Urbana has used mass spectrometry to identify the signalling peptides in individual rat pituitary and brain cells.

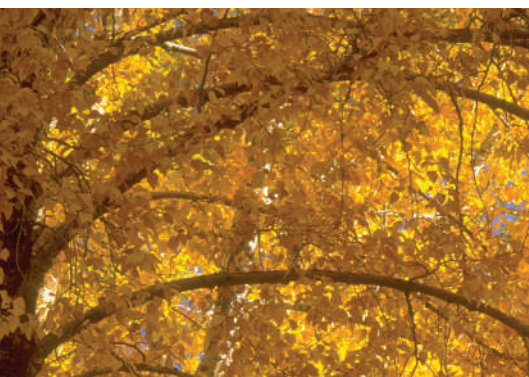
To do so, Jonathan Sweedler and his colleagues had to overcome the difficulty of working with small and fragile mammal cells. They did this by soaking the cells in glycerol, which prevented them from leaking or drying out. Applying the latest generation of very sensitive mass spectrometers to cells should allow scientists to follow changes in signalling peptides in cells isolated during specific neural processes such as learning.

PLANT BIOLOGY

Tree of knowledge

Science **313**, 1596–1604 (2006)

An international team of researchers has created the first sequence of a tree genome:



Populus trichocarpa, or the black cottonwood (pictured below, left).

Gerald Tuskan of Oak Ridge National Laboratory, Tennessee, and colleagues used shotgun techniques to sequence the genome and identify more than 45,000 putative protein-coding genes.

They compared the sequence with that of *Arabidopsis*, a herbaceous plant, and found that *Populus* has undergone a complete genome duplication since diverging from *Arabidopsis*. This has created multiple versions of genes, which can be differently expressed.

Understanding how the tree's genes differ from those of smaller plants could enable manipulation of height, trunk diameter and canopy width for commercial purposes.

QUANTUM PHYSICS

Tele for two

Nature Phys. doi:10.1038/nphys417 (2006)

Teleporting a quantum state between photons is old hat. The same feat has been achieved between ions. But Qiang Zhang of the University of Heidelberg, Germany, and his colleagues have now teleported a polarization state between two pairs of photons, a step towards creating the quantum-teleportation technology that might one day power quantum computers.

The key to the advance is a tweaked high-intensity laser that can produce enough of the bursts of six photons that the process requires. The authors reproduced various initial two-photon states in a remote receiver in between 65% and 86% of cases — comfortably above the 40% maximum

success rate that, according to the laws of quantum mechanics, could be achieved using other methods.

PALAEONTOLOGY

Early embryos

J. Paleontol. **80**, 811–825 (2006)

Tiny fossil eggs and embryos from very early multicellular animals have been identified in northwestern Canada. The find shows that the creatures lived around the globe about 540 million years ago, in the period known as the Cambrian explosion when the main animal groups first appear in the fossil record.

The millimetre-sized fossils were identified by Leanne Pyle of the Geological Survey of Canada in Sidney, British Columbia, and her colleagues after a 2003 helicopter expedition into the remote Wernecke Mountains in the Yukon. Previously, such microfossils — phosphate mineral records of cells from different species — had been found only in China and Siberia.

The team now plans to use X-ray tomography to look inside the fossils.

VIROLOGY

Follow that tag

Nature Meth. **3**, 817–824 (2006)

Watching the HIV virus move around within cells is tricky, because attaching a large fluorescent tag to its proteins can change the behaviour of the virus. Pierre Charneau at the Pasteur Institute in Paris and his colleagues found a way around this. They used a small tag that lights up when mixed

D. GULIN/CORBIS

with a fluorescent compound.

After the viral RNA is converted to DNA, the glowing protein shows how the DNA moves from the cell membrane to the wall of the nucleus and integrates into the cell's genome. The technique could be used to reveal how anti-HIV drugs might block these processes, or to examine other viruses.

NANOTECHNOLOGY

Dark secrets

Nano Lett. doi:10.1021/nl061493u (2006)

The ancient Greeks and Romans used nanoparticles to dye their hair black, say Phillippe Walter of the CNRS National Centre for Museum Research and Restoration in Paris and his colleagues.

The two-thousand-year-old recipe — which, going by the name Grecian Formula, is sold to this day as a remedy for greying hair — is a paste made from lead and water. On the hair, this reacts with sulphur in keratin protein to precipitate dark lead sulphide crystals. Walter and colleagues used X-ray diffraction and electron microscopy to show that these particles are just 5 nanometres across. Keratin fibres form a tiny matrix that contains the chemical reaction and controls the crystals' growth and arrangement.

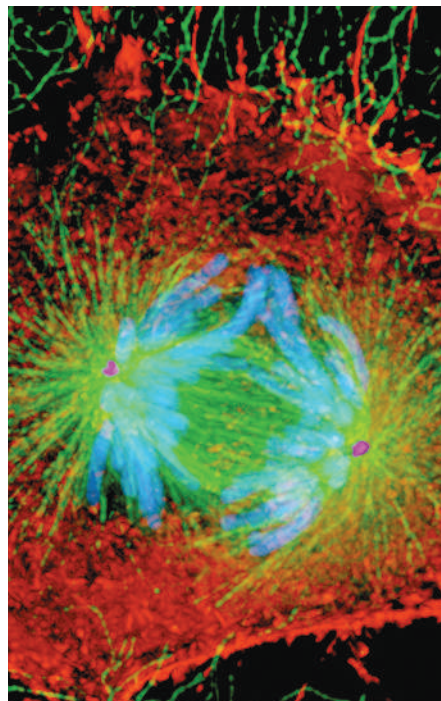
CELL BIOLOGY

Pole to pole

Nature Cell Biol. doi:10.1038/ncb1474 (2006)

When a cell divides, it hauls its chromosomes apart with a set of molecular ropes called a spindle. Researchers now report a protein that supports the two opposing ends of the spindle against the strain.

Tadashi Yamamoto and his colleagues at the University of Tokyo studied structures called



centrosomes, which seed the formation of the spindle and anchor its two ends (see picture, above). They discovered a protein, which they gave the name Kizuna, that braces the centrosome against the tension generated by the spindle as it forms. Cells lacking Kizuna could still divide, but had malformed nuclei.

COSMOLOGY

Illuminating dark energy

Astrophys. J. **648**, 884–889 (2006)

Astronomers have devised a method for studying uncertainties in measurements of dark energy, the mysterious force thought to be pulling the Universe apart.

Dark energy was discovered when studies

of the brightness of supernovae — used to measure the objects' distance — showed the expansion of the universe was accelerating. But some suspect that uncertainties in measurements of the supernovae may be distorting the effect.

Adam Riess and Mario Livio of the Space Telescope Science Institute in Baltimore, Maryland, suggest that studies of Type 1a supernovae from 9 billion to 11 billion years ago, before dark energy became dominant, could reveal how other factors, such as mass and composition, might affect the objects' brightness. But such studies will require the next generation of cameras on the Hubble Space Telescope and other space-based instruments.

CANCER BIOLOGY

Radical measures

Cancer Cell **10**, 241–252 (2006)

Cancer cells produce abnormally high levels of reactive oxygen species, which seem to sustain cancerous growth. Peng Huang from the University of Texas and his colleagues have now found a way to turn these substances against cancer cells.

Very high levels of reactive oxygen species are usually toxic to cells, so cancer cells depend on antioxidant systems that counteract these effects. Huang and colleagues found that β -phenylethyl isothiocyanate, a molecule found in vegetables such as watercress, killed cancer cells *in vitro* by disabling the cells' antioxidant systems. Normal cells were spared, probably because, the researchers conjecture, they contain a low baseline level of reactive oxygen species. When mice with cancers were treated with β -phenylethyl isothiocyanate they survived nearly twice as long as untreated mice.

JOURNAL CLUB

Lynn J. Rothschild
NASA Ames Research Center,
Stanford University, California

A microbiologist explodes the myth of the unculturables.

A surprising new trait has emerged as the darling of researchers who study bacteria and archaea. It has a binary simplicity, and the weight of references (47,000 hits on Google, 1,670 on Google Scholar) suggest that it is important. This irks me immensely.

The trait is the label 'unculturable'. There's no one paper to pin the blame on — it's widely assigned, from papers in prestigious journals such as *Applied and Environmental Microbiology* to student presentations (although any of my students using the term get berated). The fact that an organism is, as yet, uncultured is of interest to others. But since when did a human failing, the inability to culture an organism, become a trait?

Sadly, many seem to have

stopped trying to culture. The introductory remarks to many papers repeat, like a mantra, "it is well known that 99.9% of all organisms are unculturable, and thus sequencing was used to assess community diversity". Perhaps you have even written this yourself.

Now, DNA sequencing is a quick way to compare or count species. But as a technique, it is prone to contamination, and reveals little about the biology or ecology of the organisms identified.

What is to be done? I think that

a two-pronged approach is called for. In the long run, we should hold in higher esteem those with a microbial 'green thumb'. In the hands of the late greats and today's few wizards, perhaps as many as 80% of organisms are culturable.

Second, I would like papers and presentations to be honest. People should say that culturing was not successful or not attempted, not that organisms are intrinsically unculturable. It is, after all, well known that 99.9% of the time, candour is the best antiseptic.

NEWS

Claims of brain drain follow CDC reshuffle

Preventing disease has always lacked the prestige of curing it. Now advocates for public health are concerned that the field is being further undermined by dissent behind the doors of the US Centers for Disease Control and Prevention (CDC), one of the world's highest-profile public-health agencies. And many fear that if the United States is faced with a health crisis — an outbreak of pandemic flu, say, or the next SARS — the cracks will become chasms.

The federal agency, which works to combat outbreaks of infectious disease and chronic conditions such as diabetes, has seen discontent grow in the past few years over a lengthy reorganization. Concerns resurfaced earlier this month after the *Atlanta Journal-Constitution* revealed that five former CDC directors had sent a letter of concern about staff morale to the agency's director, Julie Gerberding, last December. The article sparked a record 90,000 hits and a flurry of discussion on a blog that discusses CDC internal affairs (www.cdchat.net) and in wider public-health circles. According to a crude poll on the site, 82% of voters agreed that the newspaper article accurately reflected the situation at the agency.

Established 60 years ago, the CDC has long been known for its expertise in investigating

infectious outbreaks. But after the anthrax attacks of autumn 2001, many acknowledged that the agency needed a shake-up because its 11 national centres of expertise lacked coordination, communication and efficiency.

Gerberding launched her reorganization, called the Futures Initiative, in June 2003, a year after taking office. Complaints soon began when some employees felt they were being sucked into multiple, officious committees. And when the new structure was announced in spring 2004, some were dismayed at the introduction of four coordinating centres that, critics say, created an extra layer of management between scientists and the director and stripped senior scientists of control over their budgets.

Many CDC employees are reluctant to talk openly about their concerns — and public-health experts outside the agency are also reticent because they are often closely linked with, or receive money from, the agency.

But privately, CDC employees say they are demoralized by the reorganization because it has introduced extra bureaucracy, lowered the status of science and placed too much emphasis on 'spin'. They say these changes, and the new corporate management style, are ill-suited to an agency that is supposed to investigate and



Some scientists are concerned that the US is being left poorly prepared for a health crisis.

protect public health. "The message from the current leadership is that the important scientific issues are decided elsewhere; we just have to look good to the media and not challenge conventional wisdom," says one senior public-health researcher at the CDC.

The sour situation is thought to be one reason behind a wave of high-level departures:

Health agency backs use of DDT against malaria

After decades of being shunned as an environmentally damaging chemical, the pesticide DDT is once again being touted as the most effective way to fight malaria.

The World Health Organization (WHO) announced on 15 September that it will support the indoor spraying of pesticides generally, and DDT specifically, to control mosquitoes in countries with high rates of malaria. The US Agency for International Development signalled a similar shift in policy back in May.

Although these agencies never formally opposed DDT, they did not fund countries to purchase it, and instead actively promoted the

use of insecticide-treated bednets. Malaria rates have continued to rise in the meantime, claiming more than a million lives a year, mostly in sub-Saharan Africa. The agencies now advocate combining the two approaches.

"I have to pinch myself a little to believe that they've done this, but I'm really, really happy they have," says Amir Attaran, professor of law and medicine at the University of Ottawa, Canada, who has long criticized the agencies for their malaria policies.

In sharp contrast to its previous stance, the WHO also admitted for the first time that it stopped supporting DDT despite evidence

of its effectiveness. "There are lots of data there, but people are so emotional about the issues," says Arata Kochi, director of the WHO's Global Malaria Programme. "Science comes first and we must take a position based on the science and the data."

DDT, or dichlorodiphenyl-trichloroethane, is an organochlorine that is more effective, cheaper and longer-lasting than the alternatives. Fears about its use date back to the 1960s when Rachel Carson's book, *Silent Spring*, famously chronicled its devastating effects on the environment. In the years that followed, the United States

and many European countries banned DDT.

These countries once used thousands of tonnes of the pesticide for agricultural purposes. But the use of DDT for malaria control is very different: small quantities are sprayed once or twice a year on the inside walls and ceilings of houses. Following widely publicized success with DDT in some countries such as India and South Africa, others began clamouring for the pesticide. "A lot of countries, especially in southern Africa, have become bullish about the use of DDT," says Richard Tren of the non-profit group Africa Fighting Malaria.



NEW ORLEANS CLEARED OF 'TOXIC SOUP'

Surveys show no evidence of long-term health risks caused by Katrina.

www.nature.com/news

ABD RABBO AMMAR/ABACA/EMPICS



K. LAMBERT/AP

at least eight directors of the former national centres of expertise have left since 2004. The repercussions are being felt both nationally and internationally, because the CDC plays a central role in coordinating public health across state and local health departments, as well as international responses to emerging infectious diseases. "Most people in public health are very concerned to see this level of a brain drain in the CDC," says Jeffrey Levi, head of Trust for America's Health, a non-

profit organization based in Washington DC that works to promote disease prevention.

Observers lay some of the blame on the Bush administration, which they say encouraged the agency to focus on preparation for bioterrorism at the expense of other needs. "All the emphasis was on terrorism, without willingness to recognize that the public-health infrastructure has been getting weaker for years," says Anthony Robbins, a professor of public health at Tufts University in Boston. In the 2006 financial year, the CDC received funding boosts for bioterror and pandemic-flu planning while many chronic-disease prevention programmes were cut.

The accusation that politics is usurping science has also reared its head. Critics say pressure from the administration stops the agency from investigating pressing public-health issues, such as whether abstinence-only programmes work in the fight against HIV or whether junk food is fuelling the obesity epidemic. "There is not a feeling that science drives the agenda," says a senior official who left the CDC more than five years ago. Others are critical of Gerberding herself for not resisting these political pressures and fighting for the agency's agenda.

The fear now is for what would happen if the country had to deal with a public-health crisis. Many in the field draw parallels between the CDC and the Federal Emergency Management Agency, the organization so heavily criticized over its inadequate response to Hurricane Katrina. "Our preparedness has been deteriorating in fairly dramatic and drastic ways," says Phyllis

"You can't manage a place where people don't trust you."

Freeman, who specializes in public-health policy at the University of Massachusetts, Boston.

But CDC spokesman Tom Skinner says that "there is no merit whatsoever" in the argument that the agency cannot protect public health. He says most employees understand that the agency needs to change and they have been thoroughly involved in the process. He adds that Gerberding and the agency's executive management are aware that some employees are unhappy with the reorganization and that, starting later this month, two people will be hired as ombudsmen to deal with some of their concerns. Gerberding was not available for comment.

Critics offer no clear solution for the CDC's woes, although some want Gerberding replaced with a stronger advocate for public health. "You can't manage a place where people don't trust you," says one CDC employee. Robbins says he hopes to propose a 'strike fund' for public-health workers who want to speak out about their concerns — to cover wages or legal costs for anyone who loses their job as a result.

Troubles at the CDC are symptomatic of the persistently low profile that the United States gives to public health, a field that is sometimes sidelined politically and financially when compared with research in new drugs. (Test: do you know who the US surgeon-general is?) The CDC loses out politically because it is based in Atlanta, Georgia, rather than Washington DC. And, according to Freeman, it "is pitiful in its ability to draw attention and bring funding" compared with the National Institutes of Health.

Helen Pearson



V. MOOS/CORBIS

More than a million lives are lost to malaria each year.

Even environmental groups that once vehemently opposed DDT recognize its value in malaria control. "We still think that DDT is trouble," says Ed Hopkins, director of the environmental quality programme at the Sierra Club,

a conservation group based in San Francisco. "But in some situations, where there are no alternatives, the well controlled use of DDT is better than having millions of people die from malaria."

Other groups say they are

concerned that the agencies are not dedicating enough resources to developing longer-term alternatives to DDT. "DDT is a short-sighted response with long-term consequences," says Paul Saoko, director of Physicians for Social Responsibility in Kenya.

Before countries can begin using DDT, the WHO must map resistance to pesticides to determine where spraying is likely to be effective and which pesticide would be best. Spraying won't work where mosquitoes bite and rest outdoors. And in most cases, mosquitoes — and with them, malaria — will return as soon as spraying stops, so the programmes require long-term commitment from both governments and donors.

But these practical hurdles can

be tackled, says Kochi. "So many people took the position that even though DDT and indoor residual spraying are effective, it cannot be sustained," he says. "My sense is, nothing is sustainable unless you decide to make it so. People make excuses."

Kochi, who also set up the Stop TB Initiative, is largely credited with the changes in the WHO's approach. Shortly after he took this job in October 2005, he demanded that pharmaceutical companies stop marketing single-drug artemisinin medicines, and only sell combination drugs, in order to prevent resistance.

"The breath of fresh air this man represents is just tremendous," says Attaran. "He's perfect for this job." **■**
Apoorva Mandavilli

Agency accused of 'illusion of integrity'

WASHINGTON DC

The US National Institutes of Health (NIH) has successfully switched to a new ethical culture and is spending millions on ensuring that rules about employee investments and outside activities are obeyed — according to a top agency official.

"There have been some bumps in the road, but most people would say it has gone reasonably well," says Raynard Kington, deputy director of the biomedical agency, which is based in Bethesda, Maryland.

Not everyone agrees. Kington was speaking to *Nature* two days after a hearing on 13 September in which unhappy congressmen grilled him and other agency officials. One described the NIH as an "ethical Potemkin village where a hollow system appears to provide the illusion of integrity".

"The NIH has changed its rules and that's a good thing, but they don't appear to really be doing anything to enforce the old rules against their most serious transgressors," said Republican congressman Joe Barton of Texas, chairman of the House Committee on Energy and Commerce. The committee's investigative subcommittee had summoned Kington and others to raise concerns about NIH management.

New ethics rules came into force at the agency one year ago (see *Nature* 437, 9; 2005), after a series of revelations about lucrative consultancy payments from companies collected by senior NIH scientists, often as they worked with or promoted the companies' drugs. Under pressure from Congress, NIH director Elias



Zerhouni banned employees from doing any outside consulting for pharmaceutical and biotechnology firms. The new rules also force senior scientists to limit stockholdings in biomedical companies to \$15,000. No scientists may hold shares in companies directly related to their research.

Kington told *Nature* that the agency has had "very high compliance with the new rules"

and that he hasn't had to take any disciplinary action. He adds that the agency is investing several million dollars in improving relevant databases, and that he is about to hire a chief ethics officer to monitor compliance.

Within the NIH there seems to be resigned acceptance of the rules. One senior agency scientist, who asked not to be identified, says that generally the rules "don't change most people's day-to-day life. The NIH is still a really good place to work. You can do your research."

But there has been an impact on morale. "We feel that Zerhouni didn't cover our backs," says the scientist. "Admitting we had to change the rules — it was almost like tacit agreement that there were pervasive problems with the system. And I don't think that any of us feel that was the case." NIH spokesman John Burklow counters that view. "Dr Zerhouni defended the agency and its scientists through public statements and at congressional hearings," he says.

At last week's congressional hearing, the agency was tackled about two researchers who remain on the NIH payroll despite recommendations from their institute directors that they be fired. An internal investigation found that they be fired. An internal investigation found that Trey Sunderland, an Alzheimer's researcher, accepted more than \$700,000 in consulting fees without disclosing them to his bosses, in violation of existing NIH rules. And Thomas Walsh, an infectious-disease specialist, received more

Scientists hit back at the press

In an unusual attack, more than 100 infectious-disease experts have signed an editorial criticizing a *Los Angeles Times* article that, they contend, unfairly impugned one of their colleagues.

The commentary by 109 scientists from 16 countries was published last week by *Clinical Infectious Diseases*. It attacks a 16 July article by reporter David Willman that examined two trials of antifungal agents, led by Thomas Walsh of the National Cancer Institute. In the past, Walsh has received financial support

from the drugs' makers.

The scientists say that Willman maligns Walsh by implying that he used inadequate doses of control drugs to make the new drugs seem more effective. They also dispute what they describe as Willman's implicit suggestion that subjects without fungal infections were inappropriately included in the trial. "It is our responsibility to stand up for clinical research and for a colleague whose integrity has been very unfairly attacked," says corresponding author Elias Anaissie of the University

of Arkansas for Medical Sciences in Little Rock.

Willman, a Pulitzer prizewinner who first exposed lucrative consulting contracts among senior researchers at the National Institutes of Health (see *Nature* 426, 739; 2003), is unmoved. "Anaissie *et al.* fail to cite a single inaccuracy in my work," he says. "Our coverage has brought to light matters of public importance." He notes that questions about the trials were first raised by others in the field (see, for example, *N. Engl. J. Med.* 341, 1152–1155; 1999). **M.W.**

**TEXTBOOK FREE FOR ALL**

Wiki-project aims to pool knowledge in free online texts.

www.nature.com/news

Mystery surrounds lab death



D. COOK/AP

John Agwunobi testifies at a congressional hearing to evaluate ethics reforms at the NIH.

than \$100,000 in unapproved consulting and speaking payments from 25 companies (see 'Scientists hit back at the press'). Both men have declined to comment.

The NIH doesn't have the authority to fire either of them because they are formally employed by the Commissioned Corps, a branch of the Public Health Service. To dismiss them, the corps must convene a board of inquiry. John Agwunobi, assistant secretary for health, told the subcommittee that it did so for Sunderland early this year, but added that because criminal charges are being considered against Sunderland, the board has been put on indefinite hold. Agwunobi, who oversees the corps, said that a board of inquiry was convened early this month to hear the Walsh case.

That didn't seem to satisfy his interrogators. "In both these cases, we are troubled about whether the NIH and the Commissioned Corps acted appropriately," said Republican congressman Ed Whitfield of Kentucky, subcommittee chairman. "Both the corps and the NIH seem passive, taking the minimal steps to enforce the rules that are the foundation of maintaining public trust."

Meredith Wadman

TOKYO

On 1 September, Yasuo Kawasaki, a 42-year-old assistant professor at Osaka University, was found dead in his lab after ingesting poison. The investigation into the case so far has left many questions unanswered — including whether the death was connected to a recently withdrawn article on which Kawasaki was a co-author.

That paper suggested that a type of DNA helicase called MCM2p plays an important role in DNA replication (W. Nakai *et al. J. Biol. Chem.* 10.1074/jbc.M603586200; 2006). It was published online on 12 July. But on 2 August, the journal marked it as "withdrawn".

Osaka University began an investigation on 9 August into whether the paper contained false data. In the midst of the inquiry, Kawasaki's body was found — police suspect that his death was a suicide, although they are still working on the case.

Many who knew Kawasaki have expressed surprise and shock at his death. "He was a talented young scientist," says John Diffley, an expert in DNA replication at Cancer Research UK's London Research Institute, who knew him since Kawasaki was a postdoc at Cornell University in Ithaca, New York. "His career seemed to be going very well."

Hisato Kondoh, dean of the Graduate School of Frontier Biosciences at Osaka, dismisses any suspicion that Kawasaki might have killed himself after being caught engaging in scientific misconduct. "The person who committed suicide was not involved in scientific fraud," he says. Kondoh declined to comment further on the ongoing investigation into the paper, saying only that the

results will be made public when they are complete.

But the events leading up to the withdrawal of the paper are far from clear. *Nature* has learned that Akio Sugino, head of Kawasaki's lab and corresponding author on the paper, had submitted it for publication without checking with all of his co-authors.

According to Japanese press reports, Kawasaki subsequently found that some of his data had been changed, so he asked Sugino to withdraw the article, and informed Kondoh. The university is also investigating several other related papers from the group.

Sugino was not available for comment. Robert Simoni, deputy editor of the *Journal of Biological Chemistry*, also declined *Nature's* request for information about why the paper was withdrawn, and refused to clarify how a withdrawal differs from a retraction. But scientists in the field describe Sugino as well respected and of high integrity — "a venerable old hand at DNA replication" as Diffley puts it.

Some of the details surrounding Kawasaki's death are also mysterious. Although suicide is relatively common in Japan, the cause of death tends to be hanging or gassing. Kawasaki died from ingesting sodium azide, a white, odourless solid that causes convulsions and respiratory failure within minutes. Such a method of suicide is extremely rare in Japan, according to National Police Agency statistics.

The suicide note was also unusual — rather than being

handwritten, it was printed out from a computer. And despite being addressed to his family, it was found, along with an empty container of sodium azide, in the lab where Kawasaki's body was discovered. According to the Osaka police department, the note was an explanation of Kawasaki's emotions. It did not mention the withdrawn paper

"I really hope the university gets to the bottom of this. It will be a shame if it is a whitewash."

or specific problems at work. It began: "Things at work have settled down. I want to resolve the problem."

When asked about rumours that there was intense pressure on Kawasaki before

his death, Kondoh said that measures had been in place to protect whistleblowers since the beginning of the incident. He added that there is no evidence of a connection between the apparent suicide and the withdrawn paper.

Japanese universities often respond slowly to suspicions of fraud, and aren't known for their transparency. Diffley says he now hopes Osaka will buck that trend. "I really hope the university gets to the bottom of what happened," he says. "It will be a shame if it is a whitewash." He points out that without a definitive investigation result, rumours of misconduct could blight the whole group. "The careers of many scientists will be affected."

Kiyoshi Kurokawa, president of the Science Council of Japan, agrees. "Japanese universities and institutions may not always take the right approach to resolving problems," he says. "But do they realize that the science community around the world is watching?"

Ichiko Fuyuno and David Cyranoski

ON THE RECORD

“Eris is the Greek goddess of discord and strife.”

Astronomer Mike Brown on choosing the official name of the dwarf planet formerly known as Xena.



“Fang Zhouzi and I should carry out a civilized duel to the death.”

Philosopher Li Ming suggests a novel way of solving a disagreement over the four-colour theorem.

“I hate to say we told you so, but we told you so.”

Climatologist Mark Serreze on the finding that Arctic sea ice shrank 14% between 2004 and 2005.

SCORECARD

**Academic freedom**

The Russian cabinet has replaced the word ‘Russian’ in the name of the Russian Academy of Sciences with ‘State’ — a seemingly subtle change that gives President Vladimir Putin the right to approve its choice of president.

**Chemists’ nights out**

Japanese police have launched a desperate search for three bottles of potentially deadly hydrofluoric acid, after an official from Shimonoseki Mitsui Chemicals who was carrying them one night got so drunk he couldn’t remember what he did with them.

NUMBER CRUNCH

A group of scientists have got together to defend researcher Thomas Walsh against a *Los Angeles Times* article that suggested he may have biased clinical trials towards drugs owned by the companies who were paying him (page 252).

109 is the number of authors who signed the editorial in *Clinical Infectious Diseases*.

554 is the number of potential conflicts of interest they declare.

Lawyers call for science to clear AIDS nurses in Libya

Lawyers defending six medical workers who risk execution by firing squad in Libya have called for the international scientific community to support a bid to prove the medics’ innocence. The six are charged with deliberately infecting more than 400 children with HIV at the al-Fateh Hospital in Benghazi in 1998, so far causing the deaths of at least 40 of them.

On 28 August, when the prosecution was scheduled to close its case, the Libyan prosecutor called for the five Bulgarian nurses and a Palestinian doctor to be sentenced to death. Attorneys from Lawyers Without Borders, who are handling the defence of the six, have responded by calling for the international community to request that the court order an independent scientific assessment, by international AIDS experts, of how the children became infected.

The medics were condemned to death in May 2004, but the Supreme Court quashed their convictions last December, following international protests that the first trial had been unfair. It ordered a retrial, which has run intermittently since 11 May at the Criminal

Court of Benghazi, based in Tripoli. A verdict is expected within weeks.

But the scientific community has so far shown relatively little interest in the case, says Emmanuel Altit, a member of the Paris bar and a volunteer with Lawyers Without Borders, who has in the past defended inmates at Guantanamo Bay. “We have knocked on a lot of doors, but we have not had much help; we hope this will change.”

One reason for the lack of interest, he says, is the widespread notion that the trial is a side-show, and that the “real decisions” will be made by diplomats (see page 245). Altit argues that diplomacy has so far failed to secure results, and that the medics’ release will only be secured by using scientific evidence to fight the case in the Tripoli courtroom. He hopes that exposing the “emptiness” of the prosecution case will ramp up enough international pressure to force governments to take action.

At present, the case has been sidelined by broader geopolitical interests in the opening of oil-rich Libya to international relations, says

L. LARBI/REUTERS



Five Bulgarian nurses and a Palestinian doctor stand accused of infecting hundreds of children with HIV.

Antoine Alexiev, another defence lawyer on the case. The United States decided in May to re-establish diplomatic relations with Libya. And Muammar Gaddafi, the Libyan leader, has been given red-carpet treatment at the European Union's headquarters in Brussels — without mention of the medics' situation.

First report

During the first trial, the Libyan government did ask Luc Montagnier, whose group at the Pasteur Institute in Paris discovered HIV, and Vittorio Colizzi, an AIDS researcher at Rome's Tor Vergata University, to examine the scientific evidence. The researchers carried out a genetic analysis of viruses from the infected children, and concluded that many of them were infected long before the medics set foot in Libya in March 1998. Many of the children were also infected with hepatitis B and C, suggesting that the infections were spread by poor hospital hygiene. The infections were caused by subtypes of A/G HIV-1 — a recombinant strain common in central and west Africa, known to be highly infectious.

But the court threw out the report, arguing that an investigation by Libyan doctors had reached the opposite conclusion. Montagnier believes the judgement was based at least partly

on mistranslation from English to Arabic of the term 'recombinant' — instead of referring to natural recombination of wild viruses, as intended, it was interpreted to mean genetically modified, implying human manipulation.

According to Alexiev, the decision to throw out the report removed all scientific content from the case, leaving a series of prejudgements, and confessions extracted under torture. "It's scandalous," he says. "This is a complex scientific affair, and it is impossible to judge it without a scientific basis."

Montagnier, whose efforts helped secure a retrial in the first place, says he too is upset by how events in Tripoli are progressing. "It's a rerun of the first trial," he says. "It's embarrassing politically for Gaddafi, but there is the pressure of the parents, who absolutely need to find a scapegoat. Of course this can't be the Libyans, so it falls on the medics."

The defence is scheduled to plead on 21 September, but Altit is not convinced that the science will be fairly heard. All

"It's scandalous. This is a complex affair, and it is impossible to judge it without a scientific basis."

attempts by the defence to present its arguments have been "systematically blocked", he claims, for example by switching the schedule. "The trial should be fair and equitable; until now it has been anything but."

Legally, the Montagnier/Colizzi report cannot be reinstated after having been thrown out, so the defence is pinning its hopes on persuading the court to appoint an independent science panel to produce a new report. The Tripoli court has resisted all such calls, says Alexiev. "We are hitting a wall, and that is unlikely to change before the end of this trial."

The defence is therefore resigned to probably losing the current trial, he says, and is setting its sights on the six's last chance: a final appeal in the Supreme Court, which could convene immediately after the Tripoli verdict, currently expected in November. "We need to convince the Supreme Court to nominate that international scientific assessment," he says.

"If international pressure isn't stronger before the appeal, the risk is large that they will be condemned to death," predicts Michel Taube, co-founder of Together Against the Death Penalty, a French non-governmental organization. "To avoid that outcome, diplomacy is not enough. We need international mobilization."

Only a combined pressure from lawyers and scientists as well as politicians will make a difference, agrees Altit. If the Supreme Court refuses a scientific assessment, then the international community will be able to ask: "What has it got to hide?" he says. "And if it agrees to a scientific investigation, then we will win." ■

Declan Butler



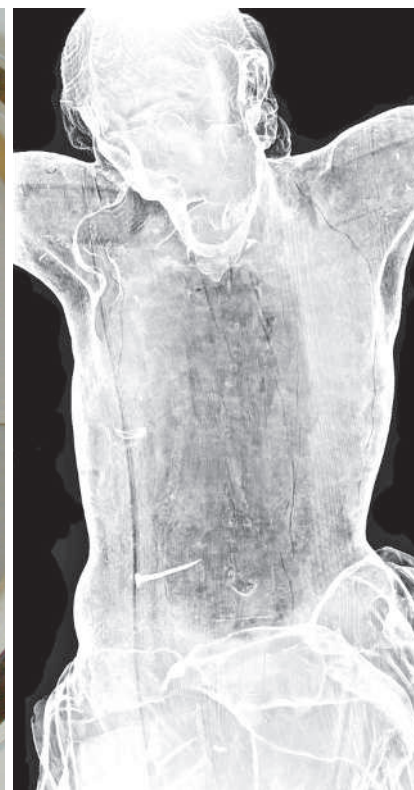
SNAPSHOT Idol scans

A precious eighteenth-century statue of Christ, *Cristo de la Peña*, that is carried in religious processions in the small town of Guadassuar, Spain, has been investigated by a novel form of imaging.

Ignasi Gironés Sarrió and Vicente Guerola Blay of the Heritage Conservation Institute in nearby Valencia are the first to use computerized multi-slice tomography on an artefact. The technique allows far more detailed three-dimensional imaging than normal computerized axial tomography. The results (pictured) show that the 1.5-metre-long figure was carved almost totally from a single piece of wood; only the arms and face were added on. Most statues of this era were made with separately carved legs.

The scan also shows that the artist reduced the weight of the sculpture by having a cavity in the torso, to make it lighter for those who had to carry it.

Alison Abbott



HERITAGE CONSERVATION INST.

That's no laser, it's a particle accelerator

Israeli physicists have turned a laser into a particle accelerator. Dubbed a *paser* — for particle acceleration by stimulated emission of radiation — the device accelerates bundles of electrons using the same principle as a laser.

At present it can only accelerate electrons by about 0.15% of their initial speed, but it could lead to compact particle accelerators and tabletop X-ray devices, according to Samer Banna of the Israel Institute of Technology in Haifa. He and his colleagues will publish their work in *Physical Review Letters*.

Conventional lasers exploit the quantum properties of atoms. An energy source is used to boost the electrons in a group of atoms into an elevated energy state. Passing light in the form of photons stimulates the atoms and causes the electrons to fall back to the lower energy level, emitting more photons in the process. These in turn stimulate more atoms and so on, so that a large number of photons are emitted. The photons are all identical, which makes the beam of light uniform.

Pasers work on a similar principle, but the output is accelerated electrons. Packets of electrons are fired into a cloud of excited carbon dioxide gas. As in a laser, the gas releases a large

number of identical photons. But those photons are instantly absorbed by the passing electrons, which get an energetic kick, and leave the device moving more quickly than when they came in.

The fact that the *paser* uses atoms to speed up electrons sets it apart from other particle accelerators. "This is unlike anything that's come before," says Eric Colby of the Stanford Linear Accelerator Center in California. The unique mode of action makes the *paser* far more efficient than current machines, which achieve acceleration by generating enormous electric fields inside huge cavities. Colby is optimistic about the *paser's* potential. "It's a pretty small effect now," he says, "but there are strong technical reasons to believe that a very significant gain in acceleration is possible."

Levi Schächter, of the Israel Institute of Technology, believes that the *paser* could also make its mark as a source of X-rays. If the high-speed electrons have their paths bent after they leave the device, they will release a laser-like beam of X-rays that could be used for medical or

nanotechnology applications. But Schächter is reluctant to guess exactly what may come of the technology. "In Hebrew we say, 'It's difficult to make predictions, particularly regarding the future.'"

It wouldn't be possible to produce the exact equivalent of a laser beam with electrons — the Pauli exclusion principle states that electrons cannot exist in the same energy state at the same time. But laser equivalents can in theory be created for other types of particle, such as gravitons (which carry the gravitational force), phonons (packets of vibration) or some nuclei — if a system can be found that emits them.

In June, for example, a group of researchers reported building a sound laser, or *saser*, that uses semiconductor technology to create a uniform beam of phonons (A. J. Kent *et al. Phys. Rev. Lett.* **96**, 215504; 2006).

Colby says that these new systems show that it is easier than one might think to generate laser-like behaviour. "If you can store energy in a material," he says, "a great many things can be done."

Geoff Brumfield

"If you can store energy in a material, a great many things can be done."

Animal-rights protesters get jail for Internet activity

A New Jersey judge last week sentenced four animal-rights activists to prison for using their website to incite harassment of the employees and shareholders of Huntingdon Life Sciences of East Millstone and the companies that do business with it.

The activists, Kevin Kjonas, Lauren Gazzola, Jacob Conroy and Joshua Harper, are members of the pressure group Stop Huntingdon Animal Cruelty (SHAC). They received prison sentences of between three and six years and were ordered to pay US\$1 million in restitution to the firm.

"People who engage in this sort of activity should know that there are real consequences to these things," says Charles McKenna, one of the government lawyers who prosecuted the case in the US District Court in Trenton.

On 2 March, a jury convicted the four SHAC members and two other members — due to be sentenced on 19 September — of interstate stalking and of conspiracy to violate the Animal Enterprise Protection Act.

Cash boost aims to spur a green revolution in Africa

Breeding new crops and training agricultural scientists will be two early aims of an initiative to stimulate farming and tackle hunger in Africa.

On 12 September, the Bill & Melinda Gates Foundation and the Rockefeller Foundation announced that they would invest an initial US\$150 million to boost

output from small farms by developing crop varieties suited to the harsh and varied African soils. It will also train crop specialists and set up systems for crop distribution, sales and advice.

The groups say they want to spark a green revolution to mirror that in Latin America and Asia in the 1940s. The project is the first investment in agricultural development by the Gates Foundation, which has previously focused on public health.

Discovery of Jupiter-lite is a massive surprise

Researchers have found an odd new planet that is nearly 1.4 times bigger than Jupiter, but only half as massive.

Dubbed HAT-P-1, the planet orbits a star much like the Sun once every 4.5 days. It was spotted by a network of automated telescopes designed to look for eclipses, or 'transits', of planets around distant stars. It is the second low-density planet spotted in recent years, and it runs contrary to current models, which predict that such objects cannot exist, says Robert Noyes of the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts. "These are very strange things," he says. The centre announced the discovery on 14 September.

'Species factory' spawns calls for protection

Diving expeditions off the northwestern coast of New Guinea have revealed Earth's richest collection of sea life, including many new species. Conservation International,



Cirrhilabrus cenderawasih, a new species of flasher wrasse found off the coast of New Guinea.

G. ALLEN

the non-profit organization leading the expeditions, is using the find to push for a dramatically broadened protected marine area in the region.

Expedition leaders are calling the area, known as the Bird's Head Seascape, a "species factory". They have found 50 previously unknown species of fish (see above), coral and mantis shrimp. The area had been home only to subsistence fishing, but scientists are worried about the increasing use of dynamite fishing and the effects of deforestation and mining on coastal waters. Currently, only 11% of the 180,000-square-kilometre reef region is protected.

Earlier this year, scientists funded by Conservation International discovered a terrestrial ecosystem with dozens of new species in the Foja mountains, just a few hundred kilometres inland from the Bird's Head Seascape (see *Nature* 439, 774; 2006).

Clean-energy institute wins promise of big bucks

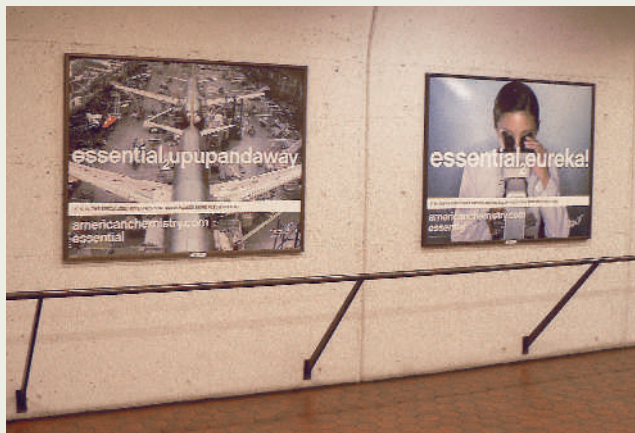
The UK government has pledged £500 million (US\$940 million) as part of its plan to set up a research institute to develop low-carbon energy sources. The centre, to be called the Energy Technologies Institute, is scheduled to open by 2008, and the government has called for the private sector to match its financial commitment.

The institute will focus on cleaner fossil fuels, renewable energy, energy efficiency and low-carbon transport and will require funding of £100 million a year over the next ten years, says a statement released by the Department of Trade and Industry last week. Power companies have until the end of November to declare an interest. Already on board are BP, EDF Energy, E.ON UK and Shell, each of which has offered £5 million a year to the scheme. "I want more companies to come forward and join us," says trade and industry secretary Alistair Darling.

US trade group rebrands chemistry

Last week, passengers on the Washington DC metro found stations decked with advertisements espousing the benefits of chemistry and the chemical industry. The campaign (see right) features pictures of fire fighters, police officers and premature babies accompanied by simple, one-line slogans such as "essential₂ living".

The adverts are part of a US\$20-million nationwide public-relations effort by the American Chemistry Council, the US chemical industry's largest lobby, based in Arlington, Virginia. The council, which is frequently criticized by environmentalists and public-interest groups (see *Nature* 432, 6; 2004), says the campaign is designed to raise public awareness of the importance of its constituent companies. "We think that our industry is not as appreciated as it should be," says council spokesman Thom Metzger.



K. SIMPSON

BUSINESS

Sequencers step up to the speed challenge

In future, hospitals should be able to sequence patient DNA swiftly and cheaply. **Rex Dalton** reports on the firms bringing that day closer.

In the summer of 1999, Jonathan Rothberg was nervously distracting himself by flipping through a computer magazine as he waited to find out whether his newborn son would survive a medical crisis. A headline about a computer chip with 44 million transistors leapt off the page; Rothberg, a chemical engineer, immediately asked himself whether some similar sort of miniaturization might one day make it easy to screen the genome of a sick child and swiftly reveal the nature of his or her illness.

Rothberg's child, Noah, is now a fit and happy seven-year-old. And Rothberg's company, 454 Life Sciences, based in Branford, Connecticut, is marketing a tabletop DNA-sequencing machine that works much faster than today's mainstream techniques. Last year, Rothberg and his colleagues sequenced 25 million base pairs (just under 1% of a human genome) in a single four-hour run¹. Today's best machines, made by Applied Biosystems of Foster City, California, work 100 times more slowly.

The 454 technique does not need to clone DNA fragments in bacterial cultures before sequencing — a stage that is necessary in the Applied Biosystems method (generally known as the Sanger method after its originator, Fred Sanger). This 'clone and pick' process is time consuming. In 454's device, every fragment of the genome being sequenced is attached to a bead in a tiny well. A pump drives reagents over the beads to create luminescent signals that show which DNA bases are being used to make copies of DNA. This yields a well-by-well set of sequences. One slide contains roughly 1.6 million wells. Running a slide's worth of reactions costs about US\$6,500 in reagents.

The 454 machines are not merely faster; the lack of a cloning stage means they can also work with more degraded material than can traditional machines. This meant the company was well placed for a headline-grabbing collaboration with the Max Planck Institute for Evolutionary Anthropology in Leipzig to sequence the genome of a Neanderthal (*Homo neanderthalensis*)².

The technology is the first sign that a new wave of companies is set to reshape the sci-

ence of gene sequencing. "454 is the first firm with commercial equipment driving change and innovation" in this area, says geneticist Edward Rubin, director of the US Joint Genome Institute in Walnut Creek, California. The institute runs some of 454's machines alongside more traditional technologies, and collaborated on the Neanderthal project. "The wild world of sequencing DNA is just beginning," says Rubin.

By some estimates, there are about 40 new ideas on how to sequence DNA percolating in labs, and several companies are exploiting them. Susan Hardin, president of VisiGen Biotechnologies in Houston, Texas, for example, has received \$8 million in US grants to develop her technique. And Massachusetts-based Helicos BioSciences plans to get its equipment into academic labs by late 2007.

Fast thinking

In the long run, the big market for these technologies is expected to be rapid analysis of a patient's genetic mutations. In the short term, all sorts of academic studies may benefit. The 454 machines have been used to do 'metagenomic' studies of ecological systems³, to unravel the genetic workings of plants⁴, and to detect rare cancer-associated mutations that are not identifiable with current sequencing methods⁵.

But the new technology is not yet in a position to dethrone Applied Biosystems. The incumbent company makes \$540 million a year selling machines that, it says, do more than 90% of today's DNA sequencing. 454 has sold just 20 machines, each worth \$500,000. It hopes, through a \$60-million marketing deal with Roche Diagnostics of Penzberg in Upper Bavaria, Germany, to place 40 more machines by the end of 2006. When asked if its Sanger-based machines will be swamped by 454's, Andrew Watson, Applied Biosystems' senior director for market development, says, "Absolutely not."

"You can't say never," he says, "But I can't see that happening for several decades. The new methods are complementary."

"The wild world of sequencing DNA is just beginning."



Step up: this machine from 454 Life Sciences has been hailed as one of the next generation of fast DNA sequencers.

A similar point was made in a recent study from the J. Craig Venter Institute in Rockville, Maryland⁶. The researchers concluded that, used together, the two technologies could improve the efficiency of whole-genome sequencing. But the 454 technology on its own was not seen as a replacement for the older technology. One crucial reason is that the individual 'reads' by the 454 technology are just 100 base pairs long; Applied Biosystems machines can do 800-base-pair reads. Especially in repetitive parts of genomes, longer reads are more easily assembled into finished sequences.

Healthy competition

But Applied Biosystems is certainly taking note of the upstarts. In July, after an eight-month survey, the firm bought Agencourt Personal Genomics of Beverly, Massachusetts, for \$120 million in order to acquire its sequencing technology. The company has also invested in VisiGen's system. "We are forcing them to respond," says 454 chief executive Chris McLeod.

"As it stands today, 454 won't take over the market," says Chad Nusbaum, co-director of the Broad Institute's sequence analysis facility in Massachusetts, which has just installed machines made by Solexa of Hayward, California. "But we could see improvements in its system that would allow it to."

One such improvement would be to increase



454 LIFE SCIENCES/PHOTOLIBRARY

the length of the individual reads. Most of the new technologies sacrifice length for parallelism and speed even more than 454 does. Early academic shakedowns suggest that Solexa's system produces chains of only 25 base pairs. Rothberg says he expects to increase 454's read lengths to 500 base pairs.

Magic number

Last month, 454's parent company CuraGen, also founded by Rothberg, hired investment bankers to study the possibility of spinning 454 off as a publicly traded firm. Such a move might provide the company with the capital to create machines that can sequence a human genome for the cost of a sophisticated test in a hospital — the idea Rothberg started with.

Might the move generate a magic \$454 million of capital? Rothberg denies the rumour that he named his company after the investment he thought it would take to reach his goal. He will say only that the name relates to a code. Maybe he'll tell the secret to Noah when he gives him his own genome sequence as an 18th birthday present. ■

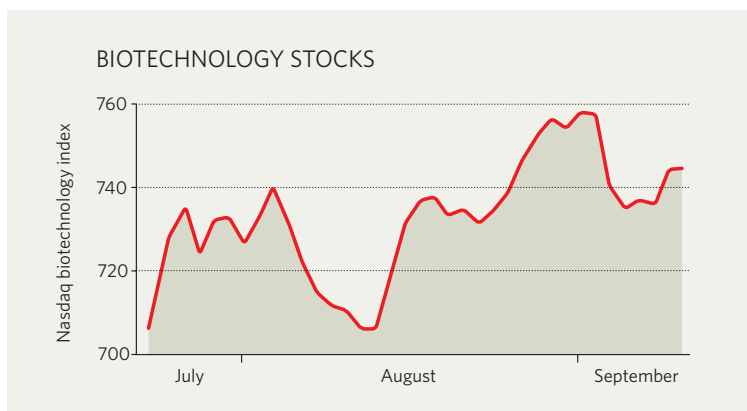
1. Margulies, M. *et al.* *Nature* **437**, 376–380 (2005).
2. Dalton, R. *Nature* **441**, 260–261 (2006).
3. Edwards, R. A. *BMC Genom.* **7**, doi:10.1186/1471-2164-7-57 (2006).
4. Henderson, I. R. *Nature Genet.* **38**, 721–725 (2006).
5. Thomas, R. K. *et al.* *Nature Med.* **12**, 852–855 (2006).
6. Goldberg, S. M. D. *et al.* *Proc. Natl Acad. Sci. USA* **103**, 11240–11245 (2006).

IN BRIEF

BEEP BEEP! IBM has won a US\$110-million contract to build the world's first petaflops supercomputer, which will run a thousand trillion computations a second. Los Alamos National Laboratory in New Mexico commissioned the machine, known as Roadrunner, to simulate nuclear explosions. When completed in 2008, Roadrunner will cut calculation times from months to weeks, according to John Hopson, Los Alamos's director for simulations. The current holder of the speed record, Blue Gene/L, is also an IBM machine — housed at rival weapons lab Lawrence Livermore in California. Hopson says Blue Gene/L's architecture makes it inefficient at running some types of weapon simulation. Roadrunner will use a more conventional architecture, achieving its speed by brute force and the use of specialized coprocessors — some of which were originally designed for PlayStation 3 games consoles.

PILL PUSHERS China's Development and Reform Commission has released plans for an invigorated drug industry. In the next five years, 5% of income from pharmaceutical sales will be reinvested in research and development. China's 5,000 or so drug companies currently invest an average of only 1% in research. This is partly because many of the firms are small, and because profit margins are low for a lot of their products — there are 300 aspirin makers in the country. Pharmaceutical companies in the United States, by comparison, typically invest 16% of revenue in research and development. The Chinese commission says it hopes to see five companies with sales of 5 billion yuan (US\$600 million) a year by 2010. Specific targets include the production of 10 to 15 new drugs and vaccines for infectious diseases, and the commercialization of 20 to 30 traditional Chinese medicines.

MARKET WATCH



This week, Wood Mackenzie, an Edinburgh-based research and consulting firm, reviews recent trends in biotechnology stocks.

The holiday season is typically quiet in the markets. Having fallen steadily since February, the Nasdaq biotechnology index flattened in July, and then managed to start climbing from mid-August. The index is up 5.7% over the past eight weeks, but still down 6.8% since the start of the year.

Good second-quarter results from most of the leading biotech companies accounted for much of this gentle growth. Smaller companies with larger gains included Neurochem from Quebec, Canada: its share value rose 88% in just under three weeks after the US Food and Drug Administration issued an 'approvable' letter for its amyloidosis treatment, Kiacta, on 11 August.

Such a letter signals the agency's readiness to approve the drug, provided it meets additional conditions — in this case more information on its efficacy. For a stock to rise on receipt of such a request is a rarity: because it is not a full marketing approval, an approvable letter almost always sends a stock down. When Encysive Pharmaceuticals of Houston, Texas, received such a letter for its heart-disease drug candidate Thelin in July, its share value fell 37% overnight.

Another large gainer this summer was Pozen of Chapel Hill, North Carolina, itself recovering from heavy losses following an approvable letter for migraine drug Trexima in June. Pozen's shares rose 27% at the end of July, when the company responded to the letter, and jumped a further 39% in early August when it announced a deal with AstraZeneca worth \$375 million. ■



SLEEP IT OFF

We've been told to eat less and move more to battle the growing obesity epidemic. But could getting more shuteye also be a way to fight the fat? **Helen Pearson** investigates.

For nutrition researcher Arne Astrup, it was a tired, overweight young girl who got him thinking. Twelve years old and 20 kilograms overweight, she was bright and more active than the average kid. Genetics and lifestyle seemed to be on her side: her slim parents packed her a nutritious lunch for school and did everything to ensure she ate healthily at home. But she was also something of a night owl. She loved to read and watch television into the night and did not sleep for more than seven hours.

Astrup was aware of emerging research pointing to links between poor sleep patterns and appetite. After he discussed the possible connection with the family, the girl stayed in bed longer, for nine or ten hours. Within one month of starting to sleep longer, she started to lose weight and her cravings for junk food dropped.

Scenarios such as this one, Astrup says, heightened his interest in the research. "I think the sleep story is really fascinating," he says. "It seems to be something that might be quite fundamental." He is now planning studies on whether sleep-deprived people eat more from a buffet.

Astrup is one of many medical researchers who have been struck by the tale of two growing epidemics and how they are intimately intertwined. Sleeping and eating are two of our most elementary drives. But in Western soci-

ety both are veering out of control: we starve ourselves of sleep and gorge ourselves on food. Now many biologists are asking whether our cavalier attitude to sleep could be at least partly responsible for our expanding bodies.

The idea is not as far fetched as it sounds, because there are plausible (if sketchy) biological mechanisms that could explain it. Overlaps are emerging between the brain mechanisms that control sleep and those that control appetite — mechanisms that may, in our evolutionary past, have been essential to fend off starvation. Many groups are now trying to figure out exactly how these circuits interconnect. The US National Institutes of Health (NIH) plans to commit as much as \$2 million to such studies this year.

A diet of sleep

In one clinical trial at the NIH in Bethesda, Maryland, sleep is being doled out like a drug. The trial will test whether tired, obese people lose weight after increasing their sleep by just one and a half hours each night. "It's a possible way to control obesity that might even be pleasurable," says obesity researcher David Allison of the University of Alabama, Birmingham. "Most people would say 'Sleep a little more? Yes I'd love to.'"

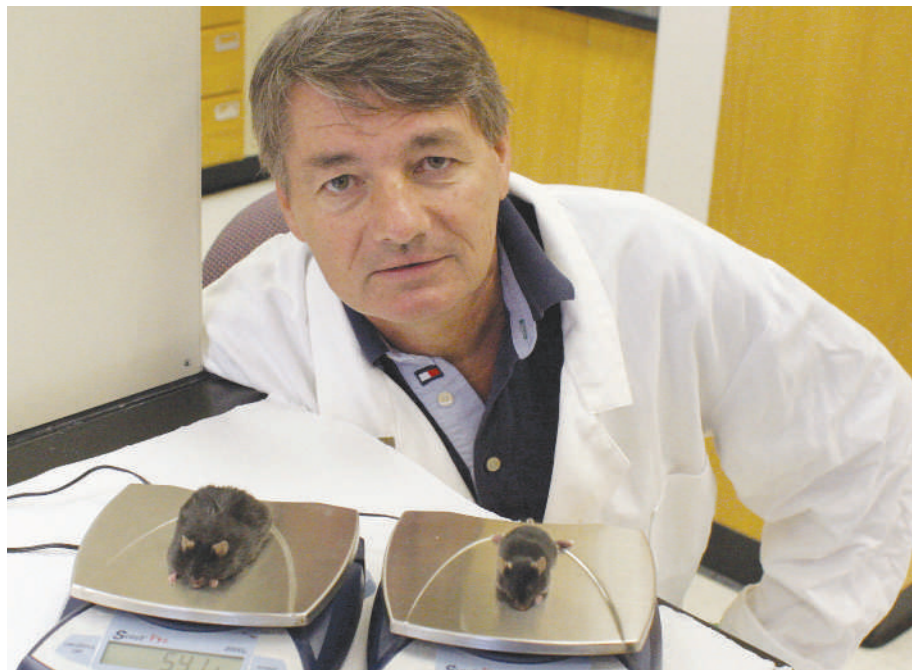
Some researchers question whether too much emphasis and money, in the obesity battle, has been placed on poor diet and lack of

exercise. They say that good sleep — an activity that we ideally spend around one-third of our lives doing — is an integral part of the package of good health. If so, then many people could be sabotaging their own attempts to lose weight simply by depriving themselves of sleep. "We need to add in a third strand," says circadian-biology researcher Fred Turek at Northwestern University in Evanston, Illinois.

It is clear that fresh ideas about obesity are needed. Two-thirds of Americans, and some 1 billion people worldwide, are overweight or clinically obese. The clinical definition of overweight is having a body mass index of 25 kg m^{-2} or more, but health risks can creep up at lower levels. The World Health Organization's 2002 World Health Report, for example, attributed about 58% of diabetes, 21% of ischaemic heart disease and 8–42% of certain cancers to a body mass index above 21 kg m^{-2} .

Researchers have pinned most of the blame for obesity on genetic make-up, too many calories and too little activity. But genes cannot, for now, be changed and efforts to get people to switch their diets and exercise habits seem to have had little effect so far.

Alongside the rising levels of obesity has been a parallel decline in the number of hours people spend sleeping — the first hint that sleep patterns could be connected to obesity. Americans' daily sleep has dropped from between



Fred Turek has found disturbed sleep patterns in mice, caused by a faulty body clock, are linked to obesity.

eight and nine hours in 1960 to less than seven hours today. A similar trend is thought to have occurred in most industrialized nations. Most people blame televisions, computers, all-hour supermarkets and the attitude held in some countries that sleep is a superfluous pastime that could more usefully be filled by school, work or play.

Burning the midnight oil

The anecdotal correlation between poor sleep and obesity has been borne out by epidemiological studies. At least a dozen reports, from different parts of the globe, in both children and adults, have consistently found that people who sleep less are more likely to have a weight problem.

In one well-publicized study¹, James Gangwisch at Columbia University in New York and his colleagues matched up the self-reported sleep habits and body mass index of more than 9,500 people from across the United States, gathered between 1982 and 1992. They found that those between the ages of 32 and 49 who slept for five hours each night were 60% more likely to be obese than those who slept for seven or more. This was true after they controlled for other obvious factors connected with obesity, such as education, age, physical activity and smoking. And a study tracking British children as they grow up showed that poor sleep in three-year-olds is an important factor in predicting obesity at the age of seven, alongside well-established risk factors such as having overweight parents or watching television for long periods².

Of course, a host of mundane explanations could account for the observation that poor sleepers tend to be overweight. One possibility is that obese people tend to sleep badly because they are obese and unhealthy. Another possibility is that people who are awake for longer simply have more time on their hands to eat. A third is that tired, irritable people lack the motivation to eat healthily or go to the gym — as Gangwisch puts it, “they might say screw this, I’ll eat a bag of chips”. For now, researchers cannot rule out some of these explanations and it is likely that they play at least a part in the connection.

But the ties between sleep and obesity are thought to run deeper than this because of another line of evidence from studies in humans and animals. This shows a biological mechanism by which sleep can promote appetite. “Some of the biggest names in the field, who people trust, are saying there is a link,” says neuroscientist Cliff Saper of Harvard Medical School in Boston, Massachusetts, who is starting his own studies into the association.

One of those big names is Eve Van Cauter at the University of Chicago in Illinois, whose laboratory has produced a series of studies showing how sleep deprivation leaves its mark on metabolism and hormones. In one study³, 12 young men were allowed to sleep for only four hours a night on two consecutive nights. The researchers tested levels of hormones including

leptin, which is released by fat cells and signals satiety, and ghrelin, which is produced by the stomach and signals hunger. They compared these hormone levels with those recorded after two nights of nine hours’ sleep.

Out of synch

After two sleep-deprived nights, the men’s blood levels of leptin had dropped by an average of 18% and their ghrelin levels had soared by 28%. At the same time, the bleary-eyed men said that they felt more hungry, particularly for carbohydrate-rich foods such as cakes, biscuits, crisps and bread, compared with proteins, fruit and vegetables.

These findings in people have been mirrored by studies in tired mice. Mimicking human sleep deprivation in mice and rats is normally difficult because the methods used to keep the animals awake, such as plunging them in water, are also stressful or require extra physical activity from the animal. So although sleep-deprived rats tend to eat far more, they still lose weight.

But studies on mice whose body clock is faulty underline the link between sleep patterns and obesity. Turek studied mice genetically engineered to lack a working version of a protein called Clock, which is active in a brain region called the suprachiasmatic nucleus that orchestrates the body’s 24-hour rhythm. This region forms part of the brain called the hypothalamus, which controls many of our unconscious body functions including sleep and appetite.

Mice without Clock lose their normal patterns of sleep, feeding and activity — and, as Turek and his colleagues found⁴, after several weeks, they grow obese and develop the hallmarks of metabolic syndrome in a few months. This has an array of symptoms including high blood sugar and cholesterol, and low insulin that, in humans, tend to be precursors to diabetes and heart disease.

lin that, in humans, tend to be precursors to diabetes and heart disease.

Together, Van Cauter and Turek’s experiments suggest that poor sleep can rapidly wreck the body’s normal systems for regulating appetite and predispose people and animals to obesity. What is not clear is the precise cellular and molecular chain of events that links one to the other — and it is these details that the NIH’s \$2-million injection is designed to reveal. “What we’re missing is the piece of evidence that would make it definitive,” Allison says.

Researchers do know that there are overlaps between the systems that regulate sleep and those that control appetite in the brain. Much

“Increasing sleep is a possible way to control obesity that might even be pleasurable.”

— David Allison

attention has focused on a huddle of cells in the hypothalamus and the two proteins they produce. These proteins are known as either orexins or hypocretins — two groups simultaneously discovered and named them in 1998. At that time, researchers found that an injection of orexins into the brains of rats stimulated feeding, indicating that the proteins were involved in controlling appetite.

Later on, other teams found that the system is defective during narcolepsy, a condition in which people fall swiftly and deeply asleep at inappropriate times. Intriguingly, narcoleptics are also often overweight. Researchers now think that orexin neurons promote wakefulness, feeding and exploration in animals.

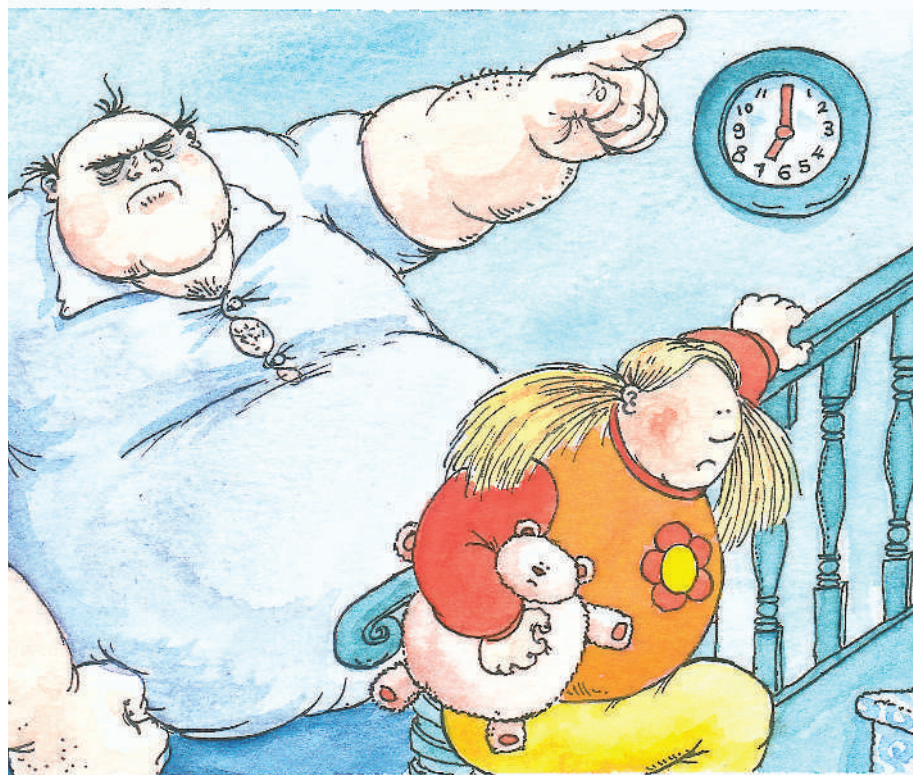
Curbing the epidemic

It is not clear how orexins are involved in the sleep-deprivation and obesity story. One idea is that lack of sleep interferes with normal circadian activity in the hypothalamus and boosts the activity of orexin neurons. This may directly affect energy expenditure and feeding, and could also alter the production of leptin, ghrelin and perhaps other appetite-control hormones. But this is just one of several possible mechanisms. "It's a concert of disharmony," Van Cauter explains. "Sleep affects our entire physiology and sleep deprivation will have adverse effects at all different levels."

In evolutionary terms, it is vital to hook up appetite and wakefulness. A mouse must eat to stay alive and be awake to forage — so it needs a mechanism that can recognize when fuel supplies are low and wake it up to find food. It is possible that humans have a similar control system that, when our ancestors were battling starvation, sent them searching for essential food. But now, anything that keeps us awake more than usual somehow taps into this circuit, sending us scavenging in the refrigerator for extra calories that we no longer burn off.

A study by Tamas Horvath and Xiao-Bing Gao at Yale University⁵ found that orexin neurons have a low threshold for activation. Overnight food deprivation boosts the formation of new synapses that excite them — which presumably encourages foraging and eating. Horvath suggests that these neurons are too easily stimulated, for example by stress or stray thoughts about work. Eating "is an unfortunate by-product of arousal," he says.

A pressing question for public health is whether the link with sleep patterns can be exploited to help battle obesity. Giovanni Cizza at the National Institute of Diabetes and Digestive



D. SIMONDS

and Kidney Diseases in Bethesda, Maryland, hopes to tackle this issue in a clinical trial funded by his institution. Cizza is recruiting 150 tired, obese people who sleep for six hours per night or less. Some of them will be taught to increase their sleep to seven-and-a-half hours (and offered a small fee as encouragement). The team will test whether, over a year, the extra sleep has an effect on weight, body fat, and leptin and ghrelin levels. If sleep can help shed even a fraction of excess weight, it could have a big impact on the health of the population, Cizza says.

"It's not like we can get everyone to sleep eight or nine hours a night and we'll solve America's obesity problem."

— Fred Turek

Following this logic, sleeping drugs could double as diet pills. Van Cauter says that she has been approached by one pharmaceutical company to discuss this possibility. Sanofi-Aventis, the Paris-based manufacturers of the widely used sleeping pill Ambien say that they do not have any planned trials. And drugs may not mimic the beneficial effect of a normal night's

sleep if they do not fully recapitulate the pattern of sleep phases.

The US Department of Defense has funded a four-year, \$2-million study to examine whether the anti-fatigue drug called modafinil, which is taken by pilots to keep them awake, could be contributing to a rise in obesity in airforce veterans. Modafinil, also prescribed for narcolepsy and certain sleep disorders, is thought to exert some of its rousing effects on the orexin system.

Many experts doubt that more sleep, be it natural or drug-induced, will be the simple answer to weight loss. Once a person is over-

weight, then poor sleep and uncontrolled appetite could become part of a vicious cycle: obesity might make it hard to sleep, and poor sleep might make weight harder to shed. "It's not like we can get everyone to sleep eight or nine hours a night and we'll solve America's obesity problem," Turek says.

Instead, researchers are keen to identify those with 'high-risk' sleeping patterns in order to prevent problems before they arise. Van Cauter is interested in finding the seemingly lucky people who survive on a few hours of sleep without appearing to suffer. (One of the many mysteries of sleep is why some people naturally sleep nine hours and others only four.) Even though this group think they can manage fine on so little sleep, they may unwittingly be vulnerable to the long-term health effects of sleep deprivation. So an early warning sign, such as altered leptin, might alert doctors that the body is actually suffering more than its owner knows.

It may, say some, be most useful to identify children who are not sleeping enough and encourage parents to change their children's sleep habits before they become ingrained. "If it's true that by sleeping half an hour extra a night our kids could be less obese it could have a huge public-health impact," Saper says.

Helen Pearson is a reporter for Nature based in New York.

1. Gangwisch, J. E., Malaspina, D., Boden-Albala, B. & Heymsfield, S. B. *Sleep* **28**, 1289–1296 (2005).
2. Reilly, J. J. et al. *Br. Med. J.* **330**, 1357 (2005).
3. Spiegel, K., Tasali, E., Penev, P. & Van Cauter, E. *Ann. Intern. Med.* **141**, 846–850 (2004).
4. Turek, F. W. et al. *Science* **308**, 1043–1045 (2005).
5. Horvath, T. L. & Gao, X.-B. *Cell Metabolism* **1**, 279–286 (2005).



A. HAAG

A testing experience

Interdisciplinary research is the new buzzword, but does a grounding in different disciplines really make you better at solving problems? **Amanda Haag** joins an experiment to find out.

Even the bastions of academia are no longer immune to reality television. In late August, 48 doctoral students arrived in the resort village of Snowbird, Utah, for a collaborative weekend of solving environmental problems — and found themselves the subject of a social-science experiment uncannily like MTV's *Real World*. First gasps and uneasy silence, then nervous laughter spread through the Cliff Lodge conference room when students learned that their hotel-suite working areas were equipped not only with laptops, Post-its, and flip charts, but also with microphones, a video camera, a voice recorder and a silent observer.

Within hours, the students were ensconced in groups of six, setting out to define a pressing environmental question, design a study and write a grant-worthy proposal to solve it — all in two and a half days. Giant sheets of scribbled notes, flow charts and diagrams soon plastered every available surface, from mini-bars to bathroom doors. Science, it seemed, could progress even with a video camera watching.

"It was a little Big Brother with the cameras and everything, but you get used to all that," says Nicole Czarnomski, a PhD student in water resources engineering at Oregon State University in Corvallis. Others initially felt a bit like lab animals. "I kept thinking, 'Don't cross

your arms, that means you're not open to an idea!'" says marine ecologist Suzanne Olyarnik of the University of California, Davis.

Being open to ideas was, after all, part of the point of spending a weekend brainstorming different approaches to environmental problem-solving. But the students were also an experiment in themselves — part of a social-science study to investigate interdisciplinary education. By sorting students according to whether they came from traditional, single-discipline graduate programmes or the new wave of interdisciplinary research, the organizers aim to learn what creative processes drive scientific thought today.

The National Science Foundation (NSF), which funds a graduate interdisciplinary programme, is not interested just in producing scientific knowledge, says Ed Hackett, a sociologist at Arizona State University in Tempe: "They're increasingly interested in understanding how it's produced."

The impact of Hurricane Katrina on the US Gulf Coast — the first anniversary of which occurred as the students set to work — is a classic example of the problems that occur when human social dynamics, natural processes and complex environmental systems collide, says Diana Rhoten of the Social Science Research Council in New York City. As a final stage of

their four-year project on the nature of scientific collaboration, she and Hackett designed the weekend as a 'charrette', a nineteenth-century concept now used to describe a short burst of intense problem-solving.

Mission: flexible

The charrette began more like a movie than a television show — with students finding packets in their working suites, "a bit like *Mission: Impossible*", says Hackett. Each packet contained not a scientific challenge, but an open-ended research design with a few defined parameters. For instance, the students might have to propose a research question that would cost roughly \$10 million over five years and support a full team of scientists. How the students defined the question would be as closely scrutinized as the way in which they went about solving it.

The questions had to deal with the intersection between human activities and ecosystem services, as defined by the 2005 Millennium Ecosystem Assessment. Each question hinged on comparing two places with different levels of human activity, such as urban New Orleans and the less developed Euphrates River in the Middle East. A suitable subject might be a region where economic growth and human health depend on fishing and a clean water supply, yet

where overpopulation threatens the aquatic ecosystem. Once the problem was defined, students worked as a group to develop the seed of a written grant proposal and give a 15-minute PowerPoint presentation to a panel of experts.

The model for the charrette dates back to the nineteenth century, when students at the architecture school of the École des Beaux-Arts in Paris were asked to draft a solution to a design problem under a strict deadline. Faculty members would often hand out problems so ambitious that few students could crack them before the cart — *charrette* — rolled by the drafting tables to retrieve the work, finished or not. Today, professions ranging from urban planning to graphic design use charrettes to inspire fresh approaches to problems.

No survivors

For this experiment, Rhoten and Hackett divided the group into eight teams of six students. Each was composed as nearly as possible of three ecologists plus one other natural scientist (such as an agronomist or a biologist), a physical scientist (such as a hydrologist or an atmospheric chemist) and a social scientist or policy analyst. There were four groups each of early- and late-stage students, divided further into students from disciplinary and interdisciplinary programmes.

That last distinction was key. One of the charrette's goals was to determine whether students trained in traditional graduate programmes work together to solve problems differently from their interdisciplinary peers. "We wanted

But did Rhoten and Hackett expect more from the interdisciplinary students because of their experience in such programmes? "I think there was a little bit of feeling at first that we're the non-IGERT people, we're from the wrong side of the tracks," says Ryan Atwell, a doctoral student in landscape ecology at Iowa State University in Ames. But Rhoten and Hackett stress that they aren't out to demonstrate the superiority of interdisciplinary programmes. "The upshot of this may not be flattering to IGERTs," Hackett says. "We have evidence from some of our fieldwork that there are some IGERT students whose disciplinary technical skills are not up to the level the discipline requires." Such problems can arise as late as when a student is submitting a dissertation. "We haven't found the right balance point between enough breadth and enough depth," he says.

Rhoten and Hackett assured the students that they were not competing with anyone, either within or between groups. And they tried to play down any *Survivor*-like aspects of the weekend. No one would be ejected from a group, although anyone could leave at any time. But the rules were complex. "It's like a Russian doll," remarked Chris Bail, a social-science doctoral student from Harvard, who served as one of the group observers. "It's got layers within the inner layers."

Students could draw only on the resources available within their group. "You don't call your faculty, you don't call your friends, and you don't ask somebody at MIT to run some data for you," Rhoten said at the opening session. To this end, each working suite contained just two laptops with Internet access and a basic software package. The researchers plan to analyse the web searches and other information students gathered during their work.

Under observation

Each group was assigned an observer, a social-science graduate student, who would sit in a corner taking notes on the group's activities. Observers were instructed not to interact with the students, and vice versa. Every 15 minutes, the observer would note exactly what was happening in the group at that moment. Was someone making a declarative statement? Was a certain person expressing scepticism? Has the group reached a low point? Have subgroups broken off? By analysing these notes, the researchers plan to look for what they call interaction hot spots, to be studied further through the video recordings.

It wasn't easy to catch up with the students, as many were tucked away in their working suites — practically around the clock. But some took time off for cocktails, hikes up the steep

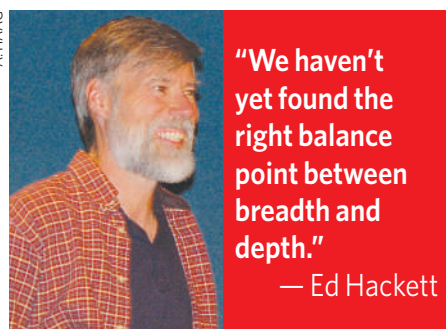
slopes of Little Cottonwood Canyon and even group trips to the hot tub. So between gulps of mountain air and after-hours refreshments, students relaxed and discussed their experiences.

All eight groups said that they gelled remarkably well. But many students expressed frustration at the initial stages, when they were bogged down in defining their research question. Some arrived at a consensus on the first morning. For others, the entire first day was filled with frustration over not being able to choose a particular ecosystem service or geographical area. "To be honest, when I went to bed Friday night, and when I got up Saturday morning, I wasn't excited about coming back," Atwell says. "I thought it was going to be one of those group experiences where you see and



"Don't call your faculty, don't call your friends, don't ask someone at MIT to run data."
— Diana Rhoten

A. HAAG



"We haven't yet found the right balance point between breadth and depth."
— Ed Hackett

to create an environment in which we could say, 'if we put these guys in teams, do they interact with individuals differently from their colleagues who haven't had this intervention?'" Rhoten explains. She and Hackett also wanted to see whether interdisciplinary programmes are training students to produce science that informs science policy. In 2005, 75 institutions across the country hosted 120 Integrative Graduate Education and Research Traineeships (IGERT), a National Science Foundation fellowship, to the tune of \$67 million and involving around 1,800 students.

respect the people you're working with, but we just don't come together." But Saturday brought a breakthrough. By then, Atwell says, "we knew who was excited about what study systems and what questions".

Semantic quibbles

When students disagreed, it was mostly over terminology. For instance, Czarnomski's group got stuck discussing what is meant by ecological resilience — a fundamental concept in ecology. "When I think about it, I think of it in more of an applied fashion and not necessarily in the theoretical constructs that a community ecologist would," Czarnomski says. Florence Bocquet, a doctoral student in atmospheric chemistry at the University of Colorado in Boulder, says that her group also quibbled over semantics. "People were explaining the same concept using their own words, phrasing the idea differently, and re-explaining the idea," she says. "I did not say, 'what you guys are saying is the same darn thing, so let's move on.' But I really wanted to."

In the end, most groups settled on proposals involving land-use change, water-resource management, and patterns of urbanization. One presentation, "The American Dream — Have a Lawn and Eat it, Too", looked at how suburban development is straining the ability of surrounding ecosystems to provide the

A. HAAG



Grad idol: a multidisciplinary panel of experts questions the students about their research proposals.

services, such as fresh water, that the communities rely on. New housing developments are being designed and marketed as having a reduced environmental footprint, although there is no comprehensive study of the true environmental impacts of individual developments. So the group proposed a comparative study between conventional and tightly grouped 'cluster' housing communities.

Horses for courses

Another group proposed comparing the Hudson River watershed in New York State and the Han River watershed in South Korea. The two basins are geographically and climatically similar, with agricultural activities and development upstream that can contaminate water downstream. The group proposed to study how human impact on the way nutrients move through the ecosystem can be taken into account in urban and rural planning.

Each team presented their findings to a seven-member panel of scientists, who ranged from pure mathematicians to ocean managers. The students pulled off relatively strong proposals, most of the experts agreed. "Were they NSF-quality proposals?" asks panel member Ann Kinzig, who studies urban ecology at Arizona State University. "Not quite yet. But did they accomplish a huge amount? Absolutely. These students only had two and a half days to work on what in some ways is a very difficult problem that senior researchers are still trying to crack"

To everyone's surprise, students tackled the

problem similarly, irrespective of whether they had interdisciplinary or traditional training. Rhoten and Hackett hesitate to draw conclusions before all the data are in, but both agreed that it was difficult to spot which groups were which. Rhoten says that the early observations bring into question whether interdisciplinary programmes train students to think or approach problems differently. "This is not a criticism of IGERTs," she says, "but it raises the question of whether they're actually changing students or just housing students who have these preferences." Although this is a long-standing question, few opportunities exist to test these hypotheses properly in group settings. Rhoten stresses that such a finding would not diminish the importance of programmes such as the IGERT. "Without the IGERT, perhaps those currently enrolled in them would have a graduate school experience that did not respond to, support or accommodate their learning preferences," she says. "That would be a huge loss."

Crossing the line

Temporary forays into interdisciplinary work, such as immersion in a charrette, may be valuable in extending the opportunity to more students without the need for multi-year fellowships. Steve Gaines, a marine ecologist from the University of California, Santa Barbara, who was not involved in the study, teaches a week-long course dealing with the interface between science and policy. "It's dramatic how

students change in just a week," he says. "So I'm not surprised at all that you can have a lot of progress over the course of a couple of days in an intensive exercise like this."

Knowing when and how to bring interdisciplinary work into one's career is a question for many researchers. Kinzig notes that many scientists feel strongly that students should become expert in one discipline before crossing boundaries. But, she adds, "I think we have an increasing number of students who aren't that interested in being disciplinary. I think if I had had to focus narrowly within a particular discipline, I would not have finished graduate school. I just would have gotten bored."

By the end of the charrette, many of the participants shared that sentiment. Ramona Walls, a doctoral student at the State University of New York, Stony Brook, was a fashion designer before she chose to become an ecologist. "It never occurred to me that I could set up a collaboration with an economist or some other social scientist, and now that I've done it, I think it's a great idea," she says.

Perhaps it is in this balance between traditional disciplinary science and synthesis-driven approaches that solutions to today's environmental and social issues will be found. "There's still glory in that other kind of science that's not problem-oriented, that's just completely curiosity-driven," Kinzig says. "We can't lose sight of that."

Amanda Haag is a freelance writer based in Colorado.



Treasure hunters: Maurice Taieb (front) at work in Ethiopia's Hadar site in 1974.

THE HISTORY MAN

Maurice Taieb laid the groundwork for the discovery of Lucy, the most famous fossil human ancestor. **Rex Dalton** meets the Tunisian-born geologist who prefers the desert to the limelight.

It was the type of desert encounter where only someone like Maurice Taieb wouldn't flinch. In the shimmering afternoon heat of the Ethiopian desert, four tribesmen materialized across the river from Taieb's geology field camp. Along with their machine guns they carried millennia of tribal animosities, the sort of hatred that can erupt in deadly skirmishes over grazing rights or livestock.

Taieb recognized the visitors as Issa tribesmen, prowling the lands of their rivals, the Afar. His Afar tribal guards and workers recognized it too, and melted away as quickly as the intruders arrived. Unarmed, Taieb turned to his few remaining colleagues and said: "Let's invite them for tea."

They sipped and bantered until the Issa left, melting back into the desert glare as silently as they had come. Taieb had manoeuvred his way through another dangerous interlude.

"Fearless" is how US palaeoanthropologist Tim White, of the University of California, Berkeley, describes Taieb, a geological legend

for his exploration of the Afar region of eastern Ethiopia. It is such a dangerous and hard place that only a handful of early-twentieth-century researchers had attempted to explore there. When Taieb started in 1966, he had the rich sediments nearly to himself. Eventually, his work would help lead palaeoanthropologists to the best hunting grounds for fossils of our earliest relatives.

Roaming the valley of the Awash River by Land Rover or donkey, Taieb picked his way through eroding gullies. Following a trail of snail fossils and encouraged by an elephant tooth, in 1968 he discovered the fossil-rich site now famous as Hadar. There, in

1974, colleagues would discover Lucy — the 3.3-million-year-old hominid that came to symbolize a new era of palaeoanthropology¹. Lucy — named after the Beatles song "Lucy in the Sky with Diamonds" but formally known as *Australopithecus afarensis* — walked upright like modern humans but still displayed primitive characteristics².

"My passion was to do research in Ethiopia, working with young Ethiopian scientists." — Maurice Taieb

In this issue of *Nature*, researchers continue Taieb's legacy with a report on a nearly complete skeleton of a three-year-old *A. afarensis*^{3,4}. From newly explored sediments just south of the Lucy site, Zeresenay Alemseged, an Ethiopian researcher at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, and his colleagues provide a stunning view of hominid evolution.

With this three-year-old *A. afarensis* female, researchers can now see in a near-complete skeleton what a mosaic of evolution the species is — the upper limbs have characteristics of gorilla or chimp, whereas the lower limbs are made for walking upright like modern humans.

Mother lode

Thanks to the Hadar site, researchers arguably have more bones of Lucy's species than of any other early hominid. At Hadar, a sedimentary layer at least 200 metres thick preserves *A. afarensis* history between about 3 million and 3.5 million years ago. The annual rains have exposed hundreds of fossils, of dozens of individual hominids. This fossil library, with today's new additions, allows researchers to do

crucial comparative analyses as with no other human ancestor.

Beyond Hadar, Taieb has helped piece together the complex geological picture of the Afar region, which yields hominid specimens back to 6 million years old⁵. With such a pedigree, he should be a big name in science. But he's barely known outside his field.

In the four decades since Taieb first went into the field, palaeoanthropologists have had their share of publicity. Some work the book and media circuit like most scientists do a library. Taieb disdains such hype-mongering; dismissing mention of headline seekers with the sweep of a hand clutching his ever-present cigarette. He seems to prefer the respect of the desert tribesmen; the headlines he wants to see are for the Africans he helped to train. "My passion was to do research in Ethiopia, working with young Ethiopian scientists so they could do the work," he says.

Taieb also stands out in a field where squabbles over credit for discoveries and permits to work at key sites are common. "He is a role model for all palaeoanthropology — an unselfish individual who invited anyone to his rich Hadar site," says geochronologist Giday Wolde-Gabriel of the Los Alamos National Laboratory in New Mexico. "He shows how people can work together for the betterment of science."

Summit meeting

And he likes getting people together. In June, the 71-year-old Taieb threw a scientific party at the CEREGE Mediterranean European Laboratory outside the French city of Marseilles, where he has been a researcher for nearly 30 years. He invited about 40 scientists to a three-day conference to reflect on Lucy's place in evolution.

It mixed backroom people with spotlight seekers, eminent Europeans with Africa's finest young rock hounds. And the Ethiopian scientists were front and centre, moderating key sessions. Only Taieb, skilled in scientific as well as desert diplomacy, could bring together such a crowd. Indeed, at a concluding dinner, Taieb was embraced simultaneously by two pillars of palaeoanthropology: White and Donald Johanson, now of Arizona State University in Tempe, who led the crew that found Lucy. Former collaborators, White and Johanson now barely speak to each other because of earlier bitter disagreements over research style and conduct.

In his lecture at CEREGE, Johanson recalled how, as a 27-year-old novice, Taieb "bubbling with enthusiasm" told him: "If you want to find fossils, you should come to the Afar." It is a place where three tectonic plates of Earth's crust come together, exposing sediments of just the right age to contain human ancestors. At



Don Johanson (left) and Tim White (centre) rarely speak, but Maurice Taieb can get them together.

the time, new dating techniques had greatly improved scientists' ability to pin down the ages of sediments and their contents. In short, it was the perfect time and place for a desert aficionado such as Taieb to start working.

Born in Tunisia, to a Tunisian father and a French mother, Taieb began his lifelong love of the African outback travelling with a merchant uncle selling goods to Bedouin on the outskirts of Tunis. He eventually made his way to France, receiving his doctorate from the University of Paris VI in 1974, with a thesis on the geology of the Awash River basin.

Decades after the French and Italians cut roads from Djibouti to Addis Ababa, Taieb followed those trails through the Afar. Again and again he circled huge areas cut by gullies and ridges. The treks gave him a vast knowledge of the landscape and its peoples; around campfires today, Afar elders still ask with interest after Maurice.

For many of Taieb's early years in the field, the biggest worries were maintaining fuel and water supplies. It wasn't until 1993 when, in what he recalls as his most danger-

ous moment, he faced down the Issa and asked them to tea. After that incident, he and his two companions — a student and an interpreter — lay away from their camp at night, ready to bolt at any sign of attack. Indeed, the area soon turned violent; a day later, a team sent by French officials to escort them out drove through a roadblock on an isolated path a few hours from Hadar. Panicked militia riddled the truck with bullets, killing an Ethiopian cultural officer and seriously injuring two others.

For anyone, the Afar was a dangerous place. The Dergue government, backed by the Soviet Union, came to power in the mid-1970s; for most of the 1980s, all field research was shut down by the political climate. Yet some of the old hands of Ethiopian anthropology think that research may

not have been as limited had certain people listened to Taieb's advice on what became known as "the grave-robbing incident".

In the late 1970s, while preparing a book about Lucy's discovery⁶, Johanson wrote about digging the femur of a modern Afar tribesman out of a rock-covered cairn, in order to compare it with a fossilized hominid knee bone later determined to be *A. afarensis*. When Taieb saw this vignette in an unpublished manuscript of the book, he says he objected vehemently, arguing that it would anger the Ethiopian authorities. His warnings were ignored. Today, Johanson will not discuss this episode. Becoming agitated when asked about these events, he denied hearing such advice.

When the book was published in 1981, jealous competitors highlighted the passage for Ethiopian officials. As predicted, they were furious. Johanson's team, on which Taieb was the geologist, and others didn't get their excavation permits back for nearly a decade.

When Johanson finally secured permits again in 1991, Taieb was dropped from the team. "He forgot to send the elevator back down," Taieb was to say years later. Some Ethiopian scientists tried to help Taieb win access to another key spot, but to no avail. He has since returned to the Afar only to help a student complete his thesis.

"I have no hard feelings," says Taieb. "It is life. I now have many letters of support from young Ethiopian scientists; this is my pleasure." And he invited Johanson to his June symposium, where the famous palaeoanthropologist spoke highly of his former geologist colleague. During cocktails before the final dinner at a historic French estate, the pair ventured off together to speak of old times. For the observing Ethiopian scientists, who had learned from Taieb in lab and field, it was another chance to mark his place in evolutionary history.

Rex Dalton writes for Nature from San Diego.

1. Johanson, D. C. & Taieb, M. *Nature* **260**, 293–297 (1976).
2. Johanson, D. C. et al. *Kirtlandia* **28**, 1–14 (1978).
3. Alemseged, Z. et al. *Nature* **443**, 296–301 (2006).
4. Wood, B. *Nature* **443**, 278–281 (2006).
5. White, T. D. et al. *Nature* **440**, 883–889 (2006).
6. Johanson, D. C. & Edey, M. *Lucy: The Beginnings of Mankind* (Simon & Schuster, 1981).



Snap happy: Taieb in 1974, with the fossil of Lucy.

Freedom of the mind got *Nature* banned by the Nazis

SIR — Today, when freedom of the press and academia are in the news, along with the neglect or misuse of scientific results and theories by politicians, it may be useful to remember a time when political demagoguery crushed these freedoms in Germany.

At the inauguration of the Philipp Lenard Institute in Heidelberg in 1936, German scientists and politicians started a campaign against *Nature* that succeeded in having the journal banned from libraries.

Nature's correspondents were accused of having created an anti-fascist espionage organization in Germany and Italy, starting in 1933. It was claimed in the science journal *Zeitschrift für die gesamte Naturwissenschaft* that *Nature* was filled with propaganda against the Nazi regime, “mostly based on democratic-liberal and Jewish feelings of hatred” (H. Rügemer *Z. ges. Naturwiss.* **3**, 475–476; 1938). Leading articles such as “Freedom of the mind” (*Nature* **139**, 941–942; 1937) and “Science and peace” (*Nature* **139**, 979–981; 1937) were listed as examples.

“When the abominable Jewish journal *Nature* speaks of the oppression of the spirit, it only means our termination, out of a sense of responsibility, of an activity by Jews and Jews-in-spirit directed at the destruction of the foundations of Aryan science in German culture,” the *Zeitschrift* article concluded. The science minister, Bernhard Rust, took action.

“In the weekly science magazine *Nature*, appearing in London, papers are often published that contain outrageous and mean attacks on German science and the National Socialist state. Therefore, this journal must be expelled from general use in scientific libraries,” he decreed in November 1937.

And expelled it was, until the end of the Second World War.

As this example shows, restricting academic freedom is a dangerous path to take. We need to resist all attempts to go down this road in the twenty-first century.

Uwe Hoßfeld*, Lennart Olsson†

*Ernst-Haeckel-Haus, Friedrich-Schiller-Universität Jena, D-07745 Jena, Germany

†Institut für Spezielle Zoologie und Evolutionsbiologie mit Phyletischem Museum, Friedrich-Schiller-Universität Jena, Erbertstraße 1, D-07745 Jena, Germany

Iran: support for science does not outweigh crimes

SIR — I was horrified to read the text of your Editorial “Revival in Iran” (*Nature* **442**, 719–720; 2006). I cannot comprehend how one of the world's leading scientific journals could publish an article calling for

scientists to adopt a benign attitude towards the present Iranian regime.

I was impressed by the sentence in the Editorial citing the fact that one of the current government's “first acts was to wipe out the debts accrued by universities”.

Nevertheless, President Mahmoud Ahmadinejad of Iran has also called publicly for my own country to be “wiped out” — from the map of the Earth. He has also been a consistent Holocaust denier, even organizing an exhibition of cartoons poking fun at the Holocaust.

At the same time, human rights are being trampled in Iran, as noted by independent organizations such as Amnesty International (see <http://web.amnesty.org/report2006/irn-summary-eng>). The regime commits many other crimes at home and abroad; it is described by the US State Department, for example, as the world's “most active state sponsor of terrorism” (see www.state.gov/r/pa/scp/2006/65559.htm).

One needs a unique degree of detachment to commend the regime for a presumably liberal attitude to science and higher education, while looking away from the dominant aspects of its essence and policies. I strongly urge *Nature* to reconsider its position.

Itamar Rabinovich

Office of the President, Tel Aviv University, Tel Aviv 69978, Israel

Several other correspondents have written to make similar points — Editor, *Nature*.

Iran is sixth, not second, in Middle East publication list

SIR — Your Editorial “Revival in Iran” (*Nature* **442**, 719–720; 2006) states that, in the 1990s, “Iran became the most scientifically productive country in the Middle East apart from Israel”. This is a misleading statement, true only when ignoring relative population size.

With a population of 68 million, Iran is the second largest country in the Middle East. It is impossible to compare small countries such as Lebanon, with 3.9 million inhabitants, or Jordan (5.9 million) to one that is 10 times more populous by merely counting the number of scientific publications. Population size should be normalized.

When the number of publications (see www.ncbi.nlm.nih.gov/entrez) is corrected for population size, Iran becomes only the sixth in terms of scientific productivity. Israel, Kuwait, Lebanon, Jordan and Saudi Arabia top the list.

Eran Meshorer

National Cancer Institute, National Institutes of Health, 41 Library Drive, Bethesda, Maryland 20892, USA

Iran's progress towards nuclear capability is no joke

SIR — One might have thought it was 1 April, rather than 17 August, on reading your indefensible celebration of Iran's nuclear programme, supposedly promoting science and education (“Revival in Iran” *Nature* **442**, 719–720; 2006). Perhaps when the fruit of this programme explodes in London, you will be writing an explanation of the humanistic ethics involved.

Guy Goodwin

University Department, Warneford Hospital, Oxford OX3 7JX, UK

Taking time to savour the rewards of slow science

SIR — As an older, experienced, part-time postdoctoral fellow, I have observed a trend amongst my younger, more vigorous colleagues to experiment themselves into oblivion. Following the lead of the ‘slow food’ movement, I suggest we adopt a philosophy of ‘slow science’ to address this issue, which I believe is damaging the very basis of scientific enquiry.

My personal choice has been to accept the here and now — I am here, making history, so why not enjoy this journey? I may not be here in six months, twelve months, two years, but I am not going to work 100 hours a week to try to attain the elusive goals of my own grant, my own lab, perhaps even tenure.

In shedding the ambition of my peers, I have discovered a secret: science, slow science, is perhaps the most rewarding and pleasurable pastime one could ever hope for. My supervisor's lab is small — two postdocs only, with no teaching responsibilities. We are free to read the literature, formulate ideas and carefully plan our experiments so as to execute thoughtful strategies. We do not plough through genomes hoping to discover something interesting; we formulate a theory, and then we go in and test it.

Perhaps we are old-fashioned, but I feel my education as a scientist has benefited far more from my five years of slow science than the preceding five years of fast science. What's more, we are on the brink of something big, exciting and wonderful, that spurs my slow science forever onwards.

Lisa Allewa

School of Biochemistry and Molecular Biology, The Australian National University, Canberra, ACT 0200, Australia

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and preferably shorter. Published contributions are edited.

BOOKS & ARTS

The road to hell?

The good intentions of a leading ecologist may not be enough to save Earth's biodiversity.

The Creation: An Appeal to Save Life on Earth

by E. O. Wilson

W. W. Norton: 2006. 160 pp. \$21.95, £13.99

Simon Conway Morris

As Anthony Burgess observed in his novel *The End of the World News* (Hutchinson, 1982), we always knew it was going to end badly. On cue and nicely resonating with our apocalyptic forebodings, we now seem to be well on course for the meltdown of civilization by environmental destruction. The threat is real, but will one more book make a scrap of difference? If it comes from E. O. Wilson, it certainly should. I fear, however, that it won't, but not for reasons of apathy or ignorance. Rather, *The Creation* fails in a much more interesting way.

How many books written by a scientist open with the phrase "Dear Pastor"? On the surface, this book is important because it is an open invitation to religious communities to bury their differences with the scientific establishment (some of whom expend considerable energy denigrating and insulting religion), join forces and so save the world from catastrophe. And who can disagree that concerted and unified action is urgently needed? This clarion call, from one of the world's leading naturalists who freely acknowledges his religious roots (even though they are now withered), must command respect. Yet despite the proffered olive branch, Wilson's thesis of cure and reconciliation is deeply problematic. Maybe Wilson's early mentors in the Southern Baptist Church were flat-Earthers. Certainly his hypothetical pastor is crippled with a fathomless biblical literalism. Such people do exist, but so too do those who grapple with the ideas of Thomas Aquinas. Significantly, not once is the pastor invited to reply: he is muzzled, perhaps the inevitable fate of a straw man.

It is a common jibe that the blame for environmental destruction should be laid at the door of reckless supernaturalists whose only concern is the next world. This thesis has long since been exploded, and in any event can be tested by exploring the spiritual foundations of the owners of factory fishing fleets, drivers of sports utility vehicles, agrochemical salesmen, property developers and those who profit from mass tourism. All of us are at fault to some degree. So other than finding common ground with religion, how is the world to be saved?



F. LANTING/MINDEN PICTURES/FLPA

Green agenda: can documenting Earth's biosphere start to reverse the decline in biodiversity?

Wilson's proposal is to embark on a massive, if not heroic, documentation of the biosphere. This, he believes, will be the catalyst to slow and ultimately reverse the relentless impoverishment of biodiversity. This seems a noble quest, but even scientifically it is intensely problematic. How can this vast inductive enterprise ever provide a coherent method of scientific conservation? More important, will it galvanize human society into collective action?

The central problem with Wilson's emergency plan is that it is ultimately a thinly disguised programme to hijack religious energy and divert it into the secular arena. Wilson briefly exposes his hand by exclaiming how this vast enterprise to catalogue the world's diversity will lead to a "transcendent and only dimly foreseeable complexity of future biology. There is to be found a new theatre of spiritual energy." Such a pantheistic agenda has not the remotest chance of working, but it also reveals a monumental misapprehension of what religion is actually trying to do.

Wilson's programme is put forward with the best of intentions, yet it is underpinned by an incoherent metaphysics. Equally important, its scientific agenda carries the real risk of imposing tyranny. Wilson is famous for his holistic programme, loosely described as 'consilience'. This aims to understand human nature in terms of entirely naturalistic processes underpinned by genetics. As part of his programme

for human development, Wilson blithely writes that one of the great goals is to "stimulate the mind with the combination of artificial intelligence and artificial emotion", chosen of course by the wisest of our leaders.

Ironically, Wilson urges us "to unglue city children from their television and computers" by reigniting their interest in the natural world. But who is to say that the very thing he deplores is not itself an inevitable outcome of evolution? Is not the rise of our technological species the next step, with the conversion of the planet to one vast farm and theme park? It is a repellent view, but I am afraid the nebulous plan offered by Wilson will not save the day.

The failure of his pantheistic agenda lies in the recurrent inability of materialists to understand that the decision to protect the biosphere can only derive from an ethical imperative that is itself independent of the natural world. Wilson is right to rage against the impoverishment of the world's biodiversity. In a striking parody of William Blake's famous lines, he fulminates against how future generations may have to repopulate a devastated world "with tigeroids ... burning artificial bright in forestoids amid insectoids that neither sting nor bite". It is a glimpse of hell, but then nobody really believes in hell, do they?

Simon Conway Morris is in the Department of Earth Sciences, Downing Street, University of Cambridge, Cambridge CB2 3EQ, UK

Teeming boisterous life

Sensuous Seas: Tales of a Marine Biologist

by Eugene H. Kaplan

Princeton University Press: 2006. 288 pp.
\$24.95, £15.95

Jon Copley

"What good men most biologists are, the tenors of the scientific world — temperamental, moody, lecherous, loud-laughing, and healthy," wrote John Steinbeck in *The Log from the Sea of Cortez* (Viking, 1951). "The true biologist deals with life, with teeming boisterous life, and learns something from it, learns that the first rule of life is living."

Steinbeck wrote those words on an expedition with his marine-biologist friend Ed Ricketts, but the description also seems to fit Eugene Kaplan, the author of *Sensuous Seas*. Kaplan's book is a compilation of entertaining anecdotes from a career that includes building fish farms in Africa, teaching science in Israel and establishing the Hofstra University Marine Laboratory in Jamaica. The book celebrates the joy of knowledge for its own sake, but also demonstrates some of the interactions between marine biology, everyday lives and even history.

Take the Mediterranean rock snail *Hexaplex trunculus*, which provided the purple dye used by royalty in the classical world. Exploiting the commercial opportunities of this snail's secretions made Phoenicia, now Lebanon, a major trading power. The same snail may also have been the 'hillazon' animal described in the Torah and the Talmud as the source of the dye used in Jewish prayer shawls. Kaplan also suggests that the fish used to feed the five thousand in the Bible would have been tilapia (pictured), whose hardiness holds promise for tackling famine through aquaculture today.

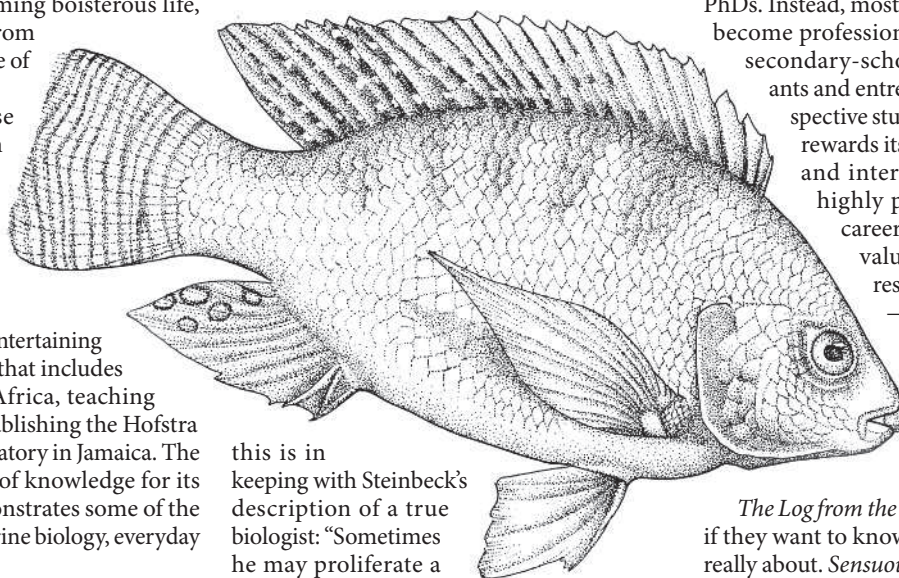
Kaplan has been teaching marine biology for half a century, and shares his experience of bringing the subject alive. Students are inspired by passion and there is plenty here, both for the subject and of the more earthy kind. Kaplan discusses with aplomb the pheromones of lamprays, the reckless mating of one-eyed shrimp, and the broken-off reproductive organs of male squid. One might conclude that he is rather preoccupied with matters reproductive, but that is true of most biologists.

There are horrors lurking here too, such as the candiru — a fish that can wriggle up the urethra of its unfortunate victim to lodge in the bladder and feast on the blood. Other perils mentioned in the book include the poisons of the pufferfish, the cone shell and the innocuous-looking fruit of the manchineel tree —

known with good reason as the 'death apple'.

This compendium of sex and death provides plenty for raconteurs wanting to leave a lasting impression on friends, relatives or — in Kaplan's case — students.

Each of the 31 chapters opens with either one of Kaplan's own memoirs or a scenario from his imagination, before exploring the marine biology behind the tale. As a result, a reader may sometimes wonder where the book is going, although the entertainment seldom flags. But



this is in keeping with Steinbeck's description of a true biologist: "Sometimes he may proliferate a little too much in all directions... meanwhile he is very good company, and at least he does not confuse a low hormone productivity with moral ethics."

Kaplan believes in hands-on teaching and advocates the necessity of literally immersing marine-biology students in their subject. Drawing on several decades of teaching field courses in the Caribbean, he describes making his students wallow in mud, chase specimens

and learn invertebrate anatomy by preparing a paella. This is not educational sadism, as students are powerfully motivated by exposing all their senses to Steinbeck's "teeming boisterous life". Watching my own students scramble across rocky shores and peer tremulously under seaweed, I can only agree that these activities ignite interest and trigger understanding.

People sometimes ask me whether the world really needs more marine-biology graduates. Despite the vast challenge of exploring and understanding the marine realm, very few students actually become career scientists, as openings are highly competitive now that universities are turning out more and more PhDs. Instead, most marine-biology students become professionals in other areas, from secondary-school teachers to accountants and entrepreneurs. I reassure prospective students that marine biology rewards its devotees with analytical and interpersonal skills that are highly prized in a multitude of careers. It also instils culturally valuable understanding and respect for the natural world — and why not have some fun along the way?

Kaplan's book conveys the breadth and excitement of an education in marine biology. I recommend

The Log from the Sea of Cortez to my tutees if they want to know what marine biology is really about. *Sensuous Seas* is now next on the list — and there is no stronger recommendation that I could make.

Jon Copley is a senior teaching fellow at the National Oceanography Centre, University of Southampton, Southampton SO14 3ZH, UK.

MORE ON THE SEA

Underwater to Get Out of the Rain: A Love Affair with the Sea

by Trevor Norton (Arrow, £6.99)

Astronomy through the years

The Cosmic Century: A History of Astrophysics and Cosmology

by Malcolm Longair

Cambridge University Press: 2006. 565 pp.
£35, \$60

Jay M. Pasachoff

John F. Kennedy once addressed what he called the "most extraordinary collection of talent, of human knowledge, that has ever been gathered at the White House, with the possible exception of when Thomas Jefferson dined alone". After diving into Malcolm Longair's epochal *The Cosmic Century*, a history of astronomy and astrophysics, he is the one I would want to dine

with if I wanted to learn about an unlimited range of material.

Longair's book is a complete reworking and major extension of a chapter he once wrote for a history book of twentieth-century physics. This background perhaps explains how such a knowledgeable astronomer was inveigled into spending the time needed to gain such a broad view. It's no surprise that the author of *Alice and the Space Telescope* (Johns Hopkins University Press, 1989) can write clearly and invitingly, but I had to sample deeply to verify that he did not give short shrift to such wide-ranging topics as the discovery of exoplanets, the use of helioseismology to analyse the inside of the Sun, and

advances in supernova research brought about by observations of SN 1987A.

Different readers may want different treatment of the subjects, depending on their mathematical backgrounds. Longair's discussion of Einstein's general theory of relativity, for example, provides sets of equations that may scare off non-physicists. But most of his treatments are non-mathematical, and many are humanized with quotations. For example, Longair not only points out that Jocelyn Bell first recognized the signals we now know as pulsars, but describes how he was present when the Soviet astrophysicist Iosif Shklovsky told her, "Miss Bell, you have made the greatest astronomical discovery of the twentieth century" (at least as far as 1970, when the encounter took place). Those familiar with the history of the discussion over credit for the discovery may note that he gives Antony Hewish, Bell's supervisor who shared the 1974 Nobel prize for physics for his work on pulsars, an acknowledgement for reading the chapter from which this book sprang.

Longair also tells the story of an underground seminar (literally, given its location in the basement of the Leiden Observatory in the Netherlands) in 1944, at which Jan Oort asked a young H. C. van der Hust to calculate whether any spectral lines would be detectable in the radio part of the spectrum. Longair's discussion describes the result (a spectral line at 21 cm) and explains how it was used to map the quasi-spiral arms of the Milky Way. He then describes the subsequent detection of 125 different molecules in the interstellar matter, and explains how the unexpected relationships of their abundances led researchers to deduce how the molecules were formed.

The section on exoplanets — unfortunately only a subsection of the chapter on the interstellar medium — explains the importance of high-precision spectroscopy, and starts with the pulsar planets discovered by Aleksander Wolszczan and Dale Frail. Longair then discusses the first optical detections of exoplanets by Michel Mayor and Didier Queloz, and a chapter endnote directs readers to *The Extrasolar Planets Encyclopaedia* if they want to keep up-to-date. I think it would have been fair to also mention the extensive work in the field by Geoff Marcy and Paul Butler, and their encyclopaedic website at www.exoplanets.org.

Little of astrophysics and cosmology escapes the gaze of Longair, a former astronomer royal for Scotland who now heads the Cavendish Laboratory in Cambridge, UK. Readers, especially those already familiar with many of the topics, will enjoy his prose. Certainly all graduate students in the field should read this book. And anyone interested in the history of science would enjoy it as bedside reading if they were willing to skip the equations. ■

Jay M. Pasachoff was president of the International Astronomical Union's Commission on Education and Development, and is at the Hopkins Observatory, Williams College, Williamstown, Massachusetts 01267-2565, USA.

Failing the ailing

Bad Medicine: Doctors Doing Harm Since Hippocrates

by David Wootton

Oxford University Press: 2006. 320 pp.

£16.99, \$25

Andrew Scull

David Wootton has made a truly remarkable discovery that he can't wait to share with the rest of us: medicine didn't work very well for millennia. Indeed, it almost certainly didn't work at all. For most of recorded history, it seems, doctors secretly did little except harm their patients while pretending to cure them. "Before 1865 all medicine was bad medicine," he writes in his book, appropriately called *Bad Medicine*. But then came Joseph Lister and antisepsis, Louis Pasteur and germ theory, and Alexander Fleming (or was it Howard Florey?) and penicillin. And from that point on, science was in place, medicine worked, and we entered an era of continuing progress and enlightenment.

Unfortunately, Wootton laments, benighted medical historians have somehow overlooked this state of affairs. Instead they have either constructed narratives of progress when medicine wasn't progressing, or (in a perverse reversal extending over three decades now) have decided that it's wrong to write tales of heroes and villains that describe an onwards-and-upwards march to scientific nirvana.

The sources of this historiographic perversity are never clearly articulated, nor are many of the offending historians identified by name, for Wootton's is a blanket condemnation. He alone has been privileged to see the true state of affairs, and henceforth the history of disease and its treatment will have to be radically recast

in the light of his fundamental new insight. For the notion that there is any sort of continuity in the medical enterprise is fundamentally mistaken, he declares: we need to develop a history of error, albeit often honest error, followed by a new history of triumph. This latter alone constitutes "the modern history of medicine, grounded in constantly developing scientific understanding".

A small fraction of the book is devoted to medicine's successes: Lister's development of antiseptic surgery, John Snow's work on cholera in the mid-nineteenth century, and the advent of antibiotics, especially penicillin. For the most part, however, Wootton's eye falls on medicine's failures and missteps. For two thousand years, the galenic and hippocratic approaches dominated Western medicine. The remedies they proffered, Wootton reminds us, were painful and actively harmful, except to the degree that they produced placebo effects. Of course, there were developments in the understanding of human biology and in medical technology in these years: Andreas Vesalius produced his stunning work on human anatomy; William Harvey discovered the circulation of the blood; Antonie van Leeuwenhoek and others used microscopes to reveal the world of the infinitely small; René Laennec invented the stethoscope; and François Magendie, Claude Bernard and others used unspeakably cruel vivisections to explore the physiology of the body. But none of these developments led to therapeutic progress, and those that might have done so (such as the ability to see germs) went unexploited for decades, even centuries.

The brief sketches Wootton provides of these and other episodes in the long reign of error are rendered deftly and vividly. When he



A saw point? Some medical practices may have done more harm than good.

attempts to address larger themes, however, things do not always go so well. Frequently he points to missed opportunities to exploit the implications of technical advances and key discoveries. He cites as examples the long hiatus between the discovery of bacteria and the link to infectious disease, and the shorter but still significant gap between the first observations of the antibacterial effects of *Penicillium* moulds in the 1870s and Florey's development of a therapy that worked in the 1940s.

But Wootton's attempts to blame these failures on the blinkered self-interest of medical

men — “doctors were determined no scientific discovery would alter their traditional therapies of bleeding, purging, and vomiting” — are crude and unsatisfying. He claims that medical historians have glorified the link between the laboratory and advances in basic science but overlooked the fact that these discoveries led to no immediate therapeutic advance. But this is simply false, as would be made clear by a quick perusal of such sources as William Bynum's *Science and the Practice of Medicine in the Nineteenth Century* (Cambridge University Press, 1994) or John Harley Warner's *The Therapeutic*

Perspective (Harvard University Press, 1986). His global characterization of the state of medical history strikes me as woefully wide of the mark — the subtle relationship between science and medicine requires a far deeper understanding than anything on offer here. And his assertion that the history of modern medicine can be reduced to a paean to scientific progress is a recipe for bad scholarship, of which there is already far too much in the world. ■

Andrew Scull is in the Department of Sociology, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0533, USA.

A miracle in sight

Adam Elsheimer painted the starry heavens in 1609.

Martin Kemp

The lynx is renowned for its sharp sight. So when the Accademia dei Lincei in Rome bound its inaugural documents into a single volume in 1603, the young German artist Adam Elsheimer provided a painted image of a lynx as its frontispiece. Founded by the Roman nobleman Federico Cesi, the new academy was dedicated to the sciences of Earth and the heavens, and provided a vital forum for Galileo and a cluster of leading intellectuals who were reforming the way that science was conducted.

Elsheimer had travelled a long way in a short time, both geographically and intellectually. The son of a tailor in Frankfurt, he travelled via Venice to Rome, where he settled by 1600. He was quickly cultivated by a circle of avant-garde thinkers and artists, including the Rubens brothers, Peter Paul (the painter) and Philipp (a humanist scholar). The former wrote emotionally on Elsheimer's early death in 1610 that “he had no equal in small figures, in landscapes, and in many other subjects”.

All Elsheimer's paintings are indeed small. They are of extraordinary visual intensity and demand unrelenting scrutiny if they are to disclose their secrets. For this reason, each visitor to the definitive exhibition of his paintings at the Edinburgh Art Festival, which closed earlier this month, was supplied with a magnifying glass. The paintings can now be seen at Dulwich Picture Gallery in London until 3 December.

The science of the Accademia dei Lincei was characterized by intense attention to visual phenomena. Within a few years of its founding, the eye was to be amplified by both the telescope and the microscope. Elsheimer's paintings declare that he was deeply immersed in this culture of taking sight as far as it would go.

Most notable in this respect is his



Flight into Egypt (shown here), characteristically painted on copper.

Within its small compass of just 31 × 41 cm, Elsheimer frames a cosmos of infinite extent, in

which heavenly and terrestrial lights sparkle in an astonishing display of observation. A band of light, the Milky Way, extends diagonally from the upper left corner. On close inspection it is not a blur of white pigment but is composed from countless tiny points of varied size (see inset). It is hard to believe that Elsheimer did not use a magnifying device; we certainly need one.

Other heavenly bodies are recognizable too. The Great Bear is visible in the upper right. The Moon, deliciously reflected in the lake, is painted not in the normal uniform manner but with a full array of surface features.

Against this display of cosmic light is set the torch of Joseph, softly illuminating the faces of mother and child as they travel on their ass to safe haven in Egypt. Further

back on the left (and partly obscured here), two peasants tend a blazing fire that shoots fusillades of sparks high into the night air.

The scene was painted in 1609, a year before Galileo published his telescopic observations. Given the circles in which Elsheimer moved, it is highly likely that he was aware of the ferment arising from the observations of the heavens occasioned by the arrival of telescopes in Rome.

Yet Elsheimer was painting a narrative, a moving rendering of a holy story. He was not simply illustrating scientific observations. His starry sky is an artful assemblage, using the new observations in the service of meaning. The Milky Way was known as ‘Jacob’s Street’ in the Middle Ages — it crossed from the ladder to heaven in Jacob’s dream. Against the lonely darkness threatening the three fugitives, the divine and immutable order of the bright heavenly bodies reassures us that Christ’s destiny will be fulfilled. Elsheimer brilliantly saw how new ways of seeing can be married to old revelations.

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK.

ALTE PINAKOTHEK, MUNICH

NEWS & VIEWS

CLIMATE CHANGE

Greenland's ice on the scales

Tavi Murray

Satellite measurements of changes in Earth's gravity field reveal ice loss from Greenland's ice sheet. Over the past four years, this melt has contributed to global sea-level rise at an accelerating rate.

The volume of the ice sheet that covers most of Greenland is so large that, were it to melt completely, sea levels across the world would rise by about 7 metres. Furthermore, an increase in its delivery of fresh water to the oceans could weaken or disrupt the 'thermohaline' circulation of oceanic salt water¹, profoundly altering the climate of the Northern Hemisphere.

Such doomsday scenarios are well rehearsed, but — expressed in this way — not necessarily accurate. If cold areas such as the centre of Greenland warm up, it might actually snow more. That would, in turn, thicken the ice sheet and remove water from the global oceans. The very different densities of snow, ice and water mean that measuring the volume of the Greenland ice sheet does not provide the complete answer as to whether it is growing or shrinking. The ideal method is to measure how the mass of the ice sheet is changing with time.

In two complementary studies, Velicogna and Wahr (on page 329 of this issue)² and Chen *et al.* (published online in *Science*)³ do just that. They show that the Greenland ice sheet lost between 192 million and 258 million tonnes of ice each year between April 2002 and April 2006 (equivalent to a volume of 212–284 km³). This rate of ice loss is equivalent to a rise in sea level of 0.5 ± 0.1 mm yr⁻¹, which is higher than many previous estimates. Both studies also show that the rate at which ice was being lost increased dramatically in the course of the study: the loss rate in the period 2004–06 was 2.5 times higher than that between 2002 and 2004 (ref. 2).

Both studies used data from the Gravity Recovery and Climate Experiment (GRACE), funded by NASA and the German Aerospace Center (DLR), which measures Earth's gravity field from space. GRACE consists of two satellites orbiting Earth. These satellites are separated by a distance of around 220 km that varies slightly as the satellites pass over anomalies in the gravity field. Since their launch in March 2002, the GRACE satellites have mapped the global gravity field every 30 days (Fig. 1). Over time, that field should show evidence of changes in the ice-sheet mass.

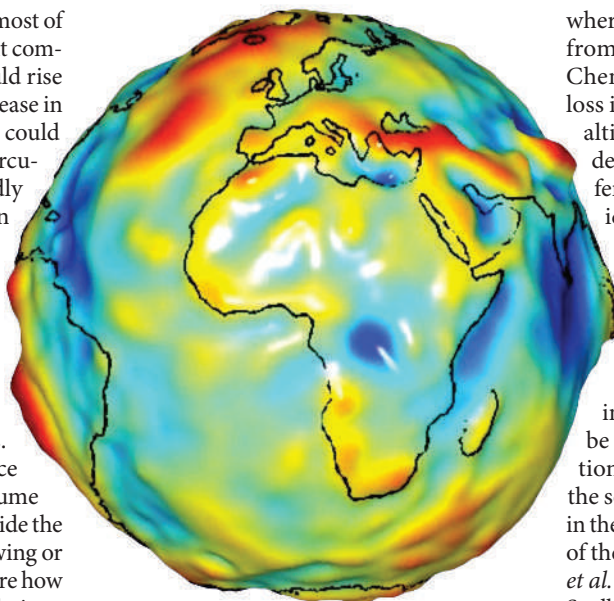


Figure 1 | Amazing GRACE. Anomaly map of Earth's gravity field, as measured by the NASA/DLR GRACE satellite.

But calculated ice-mass changes are only as accurate as the models used to remove other mass-change signals — those caused by tidal and non-tidal changes in the oceans, and by changes in the atmosphere and in Earth's mantle as it rebounds after the last ice age. At high latitudes, these models are not without error. The GRACE studies^{2,3} attempt to account for these other variations and their uncertainties, leaving a residual signal that results from the net loss of glacier ice into the oceans alone.

In particular, the high density of mantle rocks means that the gravity signal is very sensitive to even small errors in the model of rebound. But Velicogna and Wahr's estimate of uncertainty in the rebound rate² would have to be increased by a factor of ten to change their conclusion of an overall loss of ice-sheet mass to an overall gain. And as this error would be constant over the timescale of the GRACE measurements, the change in the rate of mass loss is a highly stable result.

The GRACE data can also be used to indicate

where the ice is being lost. Most of the loss is from south or southeast Greenland^{2,3}, with Chen *et al.* reporting an additional area of loss in the northeast³. Airborne and satellite altimetry over the ice sheet shows further detail of the spatial pattern, albeit over different periods. The central portions of the ice sheet, with elevations above 1,500 m, are indeed thickening, fed by increased snowfall⁴. The margins, in contrast, are thinning⁵. The outlet glaciers that feed ice from the centre of Greenland to the ocean are also depleting rapidly (Fig. 2, overleaf). This is occurring particularly in the southeast, where thinning rates can be more than 10 m yr⁻¹ (ref. 6). This depletion coincides with mass loss identified in the southeast using GRACE^{2,3}. The mass loss in the northeast² is possibly caused by thinning of the northeast Greenland ice stream⁴. Chen *et al.* postulate that changes in glaciers in the Svalbard archipelago, which lies northeast of Greenland between Norway and the North Pole, might be part of the explanation³.

The short period over which the GRACE observations were made means that the measured changes in the rate of mass loss could simply be the result of variations in snowfall or summer melt. Indeed, 2002–03, which preceded the mass-loss acceleration, was a year of unexpectedly high snowfall in southeast Greenland⁶, and 2005, which immediately followed it, was a year of record melt⁷. But the difference between snow accumulation and meltwater run-off accounts for only around a third of the mass loss from Greenland⁸. The remainder is lost through the calving of icebergs at the margins of fast-flowing outlet glaciers. The flow rate of many of Greenland's outlet glaciers increased between 1996 and 2000, and again in the period to 2005, especially in the south⁸. In spring 2004 — at the same time as the increase in mass loss recorded by GRACE² — significant accelerations in the flow and calving rate of two major outlet glaciers occurred⁹.

The agreement of the GRACE results^{2,3} with measurements of glacier dynamics⁸ on the scale and the timing of the mass loss suggests that the accelerating contribution to sea-level rise — which in 2004–06 was equal to almost

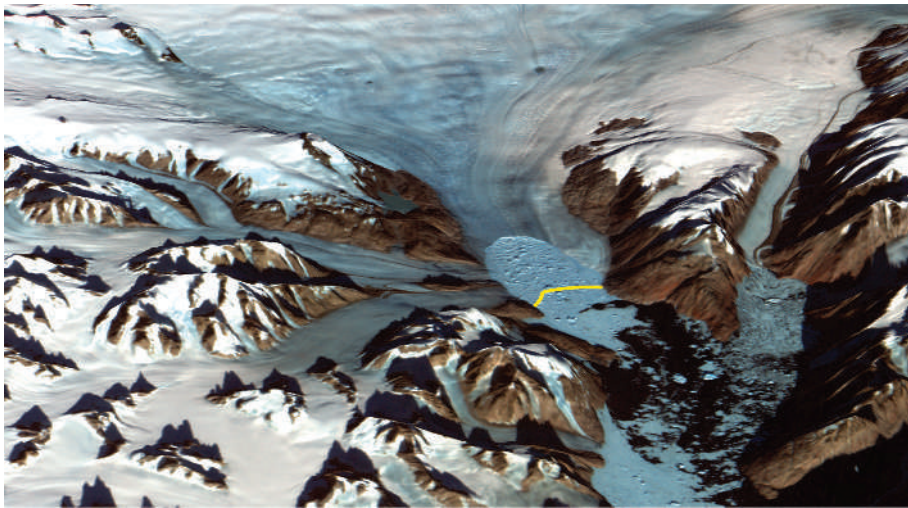


Figure 2 | Accelerated slippage. A 2005 image of Kangerdlugssuaq, an outlet glacier of the Greenland ice sheet, from a Disaster Monitoring Constellation satellite. The glacier's calving speed doubled during 2004; the yellow line shows marginal position in 2000, about 4 km in front of imaged position.

0.7 mm yr^{-1} (ref. 2) — results from changes in the dynamics of outlet glaciers. Current model predictions from the Intergovernmental Panel on Climate Change suggest a sea-level rise of $0.5 \pm 0.4 \text{ m}$ during the 21st century. But these models contain only a small component of the dynamic response of glaciers¹⁰, and the GRACE results indicate that more rapid changes are occurring than the models predict. The GRACE results can thus help us to re-evaluate the rates

of loss from the ice sheet that we should expect through climate warming.

It is clear that there is much we don't understand about the current response of the Greenland ice sheet. Records over short periods have to be treated with caution, and we cannot be certain that changes represent a profound alteration in the behaviour of the sheet. But several independent sources now confirm overall mass loss from the Greenland ice sheet,

together with unexpected and rapidly changing behaviour. Uncertainties remain, but the GRACE results provide one of the best estimates of overall mass balance of the ice sheet.

They do not, however, reveal the detailed pattern, at least not yet. It is vital that we use a variety of instruments and techniques to make continued observations of the ice sheet's response, and complement these with studies aimed at understanding the processes that are driving the observed changes. Such a programme will allow us to improve our predictive models of the Greenland ice sheet, and assess the timing and extent of its future contribution to sea-level rise.

Tavi Murray is in the School of the Environment and Society, Swansea University, Swansea SA2 8PP, UK.

e-mail: t.murray@swansea.ac.uk

1. Fichefet, T. *et al.* *Geophys. Res. Lett.* **30**, 1911 (2003).
2. Velicogna, I. & Wahr, J. *Nature* **443**, 329–331 (2006).
3. Chen, J. L., Willson, C. R. & Tapley, B. D. *Science* doi:10.1126/science.1129007 (2006).
4. Johannessen, O. M., Khvorostovsky, K., Miles, M. W. & Bobylev, L. P. *Science* **310**, 1013–1016 (2005).
5. Thomas, R., Frederick, E., Krabill, W., Manizade, S. & Martin, C. *Geophys. Res. Lett.* **33**, L10503 (2006).
6. Krabill, W. *et al.* *Geophys. Res. Lett.* **31**, L24402 (2004).
7. <http://cires.colorado.edu/science/groups/steffen/greenland/melt2005>
8. Rignot, E. & Kanagaratnam, P. *Science* **311**, 986–990 (2006).
9. Luckman, A., Murray, T., de Lange, R. & Hanna, E. *Geophys. Res. Lett.* **33**, L03503 (2006).
10. Intergovernmental Panel on Climate Change *The Science of Climate Change* (Cambridge Univ. Press, 2001).

PALAEOANTHROPOLOGY

A precious little bundle

Bernard Wood

The three-million-year old skeleton of a three-year-old child provides an outstanding resource to understand the development of a human ancestor that seems to have both walked upright and climbed through trees.

The fragile bones of infants rarely survive long enough to make it into the hominin fossil record. But if they do, they provide precious evidence about the growth and development of the individual and its species. This helps researchers not only to understand how such processes have changed during hominin evolution, but also to interpret the function and the taxonomic significance of the better-sampled adult specimens. In this respect, the remarkably complete 3.3-million-year-old skeleton of a three-year-old *Australopithecus afarensis* female, found in Dikika, Ethiopia, is a veritable mine of information about a crucial stage in human evolutionary history. In this issue, the fossil is described by Alemseged *et al.* (page 296)¹, and its geological and palaeontological context is reported by Wynn *et al.* (page 332)².

Thanks to efforts in Ethiopia and elsewhere, we already know a good deal about *A. afarensis*. It has been called an 'archaic' hominin for

at least two reasons. First, it is old: its fossils date from between 4 million and 3 million years ago. Second, its morphology is archaic, in the sense that its brain case, jaws and limb bones are much more ape-like than those of later taxa that are rightly included in our own genus, *Homo*. When adjusted for its body size, the brain of *A. afarensis* is not much larger than that of a chimpanzee, and although it has lost the large canines that distinguish apes from hominins, other aspects of its dentition, such as its relatively large chewing teeth, are still primitive (Fig. 1).

There remains a great deal of controversy regarding the posture and locomotion of *A. afarensis*. Most researchers accept that it could stand upright and walk on two feet, but whether it could climb up and move through trees is still disputed. Some suggest that its adaptations to walking on two feet preclude any significant arboreal locomotion, and interpret any limb

features that support such locomotion as evolutionary baggage without any useful function³. Others suggest that a primitive limb morphology would not have persisted unless it served a purpose⁴.

The Dikika infant is not the first early hominin infant to be found. That distinction belongs to the Taung child, whose discovery was reported just over 80 years ago⁵. What makes the Dikika infant remarkable is its unprecedented completeness for such a geologically ancient specimen. The infant was found in sediments that formed the bottom of a small channel close to where a river discharged into a lake². This was not a turbulent stream or river. The flow was sluggish, typical of the type of braided streams that make up a river delta. The corpse of the infant was buried more or less intact, and the sediment in flood waters must have swiftly covered it.

Some parts of the specimen — the pelvis, the lowest part of the back and parts of the limbs — are still missing, but what is preserved is remarkably complete. The face, the brain case and the base of the cranium, the lower jaw, all but two of the teeth (including unerupted adult teeth still in the jaw), both collar bones, the vertebrae down to the lower back, many ribs, both knee caps and the delicate bone that holds open the throat, the hyoid, are all there. Even the medial epicondyle of the humerus has survived. This is the bony projection on

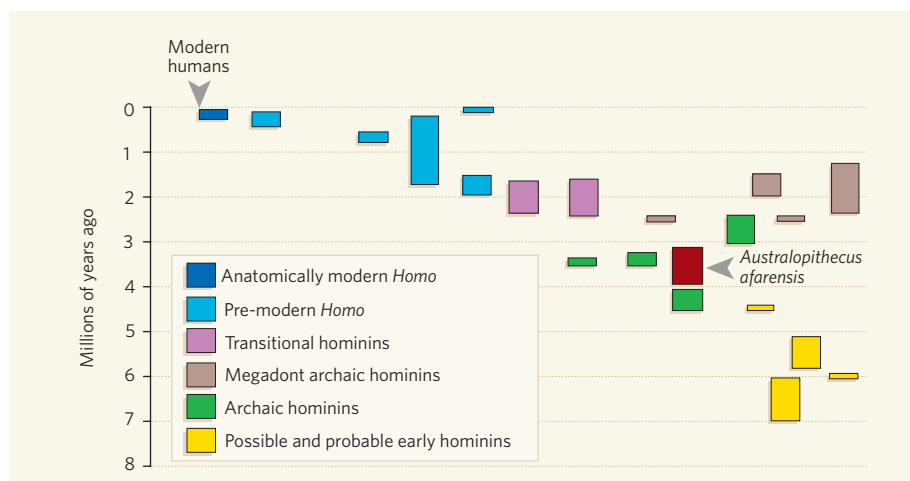


Figure 1 | A hominin taxonomy. Species are ordered according to the period of their fossil record and, left to right, according to their resemblance to modern humans: those with large brains, small chewing teeth and jaws similar to those of *Homo sapiens* are found to the left, those with large chewing teeth and jaws to the right. *Australopithecus afarensis*, an infant female specimen of which has been found in Dikika, Ethiopia^{1,2}, lived between 4 million and 3 million years ago. Its small brain is not much larger than that of a chimpanzee, but its dentition has features akin to those found in more modern hominins.

the inside of your elbow against which your left thumb rubs if you hold your right elbow with your left hand. In a three-year-old infant, this tiny piece of bone is still separate from the main shaft of the humerus. One must travel forward in time more than three million years, to a Neanderthal infant from Dederiyeh⁶, Syria, to find a comparably complete hominin infant skeleton.

This anatomical cornucopia was not evident when the specimen was found in 2000: most of the Dikika infant was invisible, hidden within a slab of sandstone. Zeresenay Alemseged has devoted many thousands of hours over a five-year period to removing, painstakingly, the cement-like matrix that surrounds the delicate bones. The patience, time, skill and effort required to preserve and expose the morphology of this and other similar early hominin fossils⁷ should not be underestimated.

But why are Alemseged *et al.*¹ so sure that the infant belongs to *A. afarensis*, and can we have confidence in its age — both the geological age of the fossil and the age of the child it represents? The geological age is secure. The Dikika sediments contain crucial evidence of the same layers of ash that have provided reliable argon–argon isotope ages at other East African fossil sites². There are also subtle and not so subtle differences between the faces of *A. afarensis* and the other hominin taxa known from similarly aged rocks, and the Dikika infant already shows signs of the type of upper jaw and nasal morphology that is seen only in *A. afarensis*. These signs are a rounded area above the upper teeth; a separation between the bone covering the roots of the upper canine teeth and the edge of the opening for the nose; and hourglass-shaped nasal bones that fit into a recess in the frontal bone much like a tenon fits into a mortise.

The second of the two age estimates, the chronological age of the infant, is less secure.

All one can do is use the kind of computed-tomography imaging familiar from modern hospitals to compare the development of the yet-to-emerge permanent tooth germs of the Dikika infant with the teeth of modern human and chimpanzee infants of known ages⁸. The best match is with three-year-old chimpanzees. But it is highly unlikely that the pace of development of *A. afarensis* was exactly the same as that of modern chimpanzees. So, for now, the chronological age of the Dikika infant must remain an informed guess.

The discoverers of the Dikika fossil have only just begun the task of capturing all the data contained in the specimen, but already these preliminary data¹ are informing the controversy of how *A. afarensis* moved. If its mode of locomotion was exclusively on two legs, one would expect that the limb bones and the organs that help it to balance would be more similar to those of the only living bipedal higher primate (that is, us) than to those of chimpanzees and gorillas. These primates walk on two feet only rarely, if at all.

Alemseged *et al.*¹ pay careful attention to the shoulder, hand and the semicircular canals of the inner ear, the morphologies of which record the motion of the body. The shoulder-bone (scapula) of the fossil is more like that of a gorilla than a modern human, and the bones of the only complete finger are curved like those of a chimpanzee. Chimpanzee finger bones begin life only slightly curved, but become more curved when the hands are used to climb branches⁹; this is what seems to have happened in the case of the Dikika infant. Lastly, images of the inner ear of the specimen show it to have semicircular canals more like those of chimpanzees than of modern humans¹⁰. The fluid-filled semicircular canals are crucial in maintaining balance, and so all three lines of evidence suggest that the locomotion of *A. afarensis* was unlikely to



50 YEARS AGO

If Pierre Charron in his "Treatise on Wisdom" was himself wise, the true science and study of man is man. Things, of course, were easier in the sixteenth century, when fossil men were not in the laboratory or the study... Alas, in recent years the study of man has been attempted and magnified by all classes and conditions of men: geologists and palaeontologists; anatomists and anthropologists; statisticians and geneticists; blood-group specialists and geochronologists; and adventurers and plain unvarnished liars.

From *Nature* 22 September 1956.

100 YEARS AGO

The recent correspondence on the subject of radium, started in the *Times* by Lord Kelvin, has...apparently closed without any very definite conclusion being reached... Lord Kelvin's opening challenge was broad and sweeping. He took exception to the statement...that the production of helium from radium has established the fact of the gradual evolution of one element into others, and denied that this discovery affected the atomic doctrine any more than the original discovery of helium in cleveite. The obvious conclusion was that both cleveite and radium contained helium. He also stated that there was no experimental foundation for the hypothesis that the heat of the sun was due to radium, and ascribed it to gravitation... Prof. Armstrong, it is true, immediately enrolled under Lord Kelvin's banner... [His] letter merely served to provide Sir Oliver Lodge with justification for his favourite theme, which appears to be that whereas chemists have an instinct of their own for arriving at their results, reason is the monopoly of the physicist, whose results the chemist usually manages to absorb in the end. No better argument against the unfairness of this could be provided than by the history of radio-activity itself, which owes at least as much to the chemist as to the physicist.

From *Nature* 20 September 1906.

50 & 100 YEARS AGO

have been restricted to walking on two feet.

I am especially intrigued by the detailed morphology of the hyoid bone in the throat of the fossil. Does the open space in the body of the hyoid mean that *A. afarensis* had air sacs in its neck? In the absence of large canines, these air sacs might have been a way in which males established a dominance hierarchy, and females judged the quality of a potential mate.

Whatever the answers to such questions, the Dikika infant has the potential to provide a wealth of information about the growth and development, function and taxonomy of *A. afarensis*.

Bernard Wood is in the Department of

Anthropology, George Washington University,
2110 G Street NW, Washington DC 20052, USA.
e-mail: bernardawood@gmail.com

1. Alemseged, Z. *et al. Nature* **443**, 296–301 (2006).
2. Wynn, J. G. *et al. Nature* **443**, 332–336 (2006).
3. Latimer, B., Ohman, J. C. & Lovejoy, C. O. *Am. J. Phys. Anthropol.* **74**, 155–175 (1987).
4. Stern, J. T. *Evol. Anthropol.* **9**, 113–133 (2000).
5. Dart, R. A. *Nature* **115**, 195–199 (1925).
6. Akazawa, T., Muhesen, S., Dodo, O., Dodo, Y. & Mizoguchi, Y. *Nature* **377**, 585–586 (1998).
7. White, T. D., Suwa, G. & Asfaw, B. *Nature* **375**, 88 (1995).
8. Dean, C. & Wood, B. A. *Digital Arch. Hum. Paleobiol.* **3**, (2003).
9. Richmond, B. G. *Am. J. Phys. Anthropol.* **120**, Suppl. 36, 178–179 (2003).
10. Spoor, F., Wood, B. & Zonneveld, F. *Nature* **369**, 645–648 (1994).

will be found because its location will eventually fall under the spreading gaussian peak. But this is not so. Even in one dimension, some particle trajectories move to the right while a hidden, static target is on the left. Averaging over all possible trajectories presents the startling fact that the mean first-passage time to any particular target is infinite.

But what if there are many targets? Consider a random walker starting on a line of targets on a square lattice. Here, the walker gradually diffuses away from the line, so the targets are visited less and less over time. One can calculate that after a large number of jumps N , only $N^{1/2}/\ln N$ of the targets are visited. In three dimensions, the result is worse: even more targets go unvisited, with only $\ln N$ sites discovered⁴.

Despite this unpromising prognosis, Bénichou and colleagues² rely, in part, on diffusion in their two-dimensional search model for non-revisitable static targets. They propose a two-state search pattern. In the first, dynamic case, the searching is diffusive, and the target is found immediately when it is within a certain distance, A . In a second, static case, the seeker is stationary and 'reacts' with the target at a certain rate when the target comes within a certain range.

In both cases, the time spent on the search phase is a random variable. When the search phase ends, either successfully or unsuccessfully, the motion of the seeker changes to a relocating ballistic: it shoots off in a random direction for a random stretch of time during which, according to the model, discovery of a target is not possible. This phase is followed by a further search phase, and so the cycle continues.

The authors derive² an equation for the distribution of first-passage times required for the seeker to find the target, for a general model combining the static and dynamic cases and allowing for an arbitrary rate of diffusion, D , and rate of reaction, k . As an

MATHEMATICAL PHYSICS

Search research

Michael F. Shlesinger

How does one best search for non-replenishable targets at unknown positions? An optimized search strategy could be applied to situations as diverse as animal foraging and time-sensitive rescue missions.

Operations research — the field that uses mathematical methods to optimize complex real-world structures and processes — grew out of the analysis of military problems during the Second World War. One such question was the optimization problem 'How to hunt a submarine'¹, the analysis of which had to take several factors into account. For instance, a figure-of-eight search pattern of an aircraft scouring littoral waters would be different from the pattern for a search of deep-ocean waters. The problem was also complicated by the fact that a negative search might mean only that a submarine was submerged, not that it was absent.

The aim of such searches is to remove the target, and, writing in *Physical Review E*, Bénichou and colleagues² bend their minds to minimizing the amount of time needed when searching for such 'non-revisitable' targets. Since the early days of aircraft hunting submarines, many types of search have been investigated in which the target may or may not be destroyed upon contact. One such study³ interpreted data from transmitters attached to the legs of itinerant albatrosses. The radio signal was silenced when an albatross was in the water. What emerged was a fractal on-off signal pattern that was consistent with the bird flying a 'self-similar' pattern — one in which the whole has the same shape as smaller component parts — with the end of each segment punctuated by a water landing. Presumably, the flight pattern reflected a search for food, and perhaps also the lifetime of thermals on which the seabirds ride.

Whatever the exact reason, this research led to several papers demonstrating that fractal patterns of the albatross type, called Lévy

flights, are an optimal search strategy. But under what conditions? The theoretical Lévy strategy has a wide distribution of flight segment lengths, and the mean of this distribution is infinite. But in the case of the albatross, the act of catching food cannot be accomplished during a flight (Fig. 1), so too much time is spent in the flight segments for a Lévy strategy to minimize search time.

At first glance, diffusion might seem to be a satisfactory strategy for minimizing search time. When a particle starting at an arbitrary origin moves randomly, its location is described by a gaussian probability distribution that spreads out with a variance that grows linearly with time. It might seem that any target



Figure 1 | Gotcha! An albatross completes a search for an non-replenishable target. Bénichou and colleagues² consider optimal strategies for such searches.

D. HADDEN/ARDEA

exact solution is not available, approximations were necessary until an equation with an analytical solution could be derived.

The approximate formula that emerged agreed well with simulations. In the case of a static seeker, abandoning a search too quickly (within time $t < 1/k$) will probably result in missing a target that is present. But spending too much time searching, $t \gg 1/k$, when a target is not present will not be conducive to optimizing the search efficiency either. Similarly, spending too much time diffusing in an area without targets is unproductive, as is spending too little time in a target-rich environment. Optimal mean times for search and relocation are possible.

Bénichou *et al.* consider a low-density system in which the average distance, B , between two targets is much greater than the distance A from which a target can be discovered. In addition, they stipulate the condition $D/v \ll A$, where v is the velocity of the seeker in the

ballistic phase. Essentially, this is a requirement that ballistic relocation should be very efficient relative to diffusion. The optimal mean search time is found to be different in the static case, where it depends on k and is proportional to $1/v^{1/2}$, and in the dynamic case, where it depends on D and is proportional to $1/v^2$. Surprisingly, however, for an optimal search in both the static and dynamic cases, the same condition is found for the mean time spent in the ballistic motion phase. This is a function of A and B , proportional to $1/v$, and independent of D and k .

In any probability-based search problem, adding information provides values for conditional probabilities. Bénichou and colleagues' analysis holds under strict conditions, especially that the seeker has no knowledge about where the perishable targets are. As soon as that condition changes, the search optimization strategy should also change. For instance, a methodical search would be best if the target must be found

regardless of the amount of time involved.

These results for non-replenishable targets might be useful for situations ranging from animals foraging for food to proteins binding on DNA. For other cases — where the target can move or trick the seeker, for example — other optimal strategies are expected. The same would be true if different cost penalties, other than minimizing the search time, are chosen: if, for instance, the ballistic phase were more costly per unit time than the diffusive phase. ■

Michael F. Shlesinger is at the Office of Naval Research, 875 North Randolph Street, Arlington, Virginia 22203, USA.

e-mail: shlesim@onr.navy.mil

1. Morse, P. M. & Kimball, G. E. reprinted in *The World of Mathematics* Vol. 4 (ed. Newman, J. R.) 2160–2178 (Simon & Schuster, New York, 1956).
2. Bénichou, O., Loverdo, C., Moreau, M. & Voituriez, R. *Phys. Rev. E* **74**, 020102 (2006).
3. Viswanathan, G. M. *et al.* *Nature* **381**, 413–415 (1996).
4. Weiss, G. H. & Shlesinger, M. F. *J. Stat. Phys.* **27**, 355–363 (1982).

HIV

Tired T cells turn around

Sarah Rowland-Jones and Tao Dong

HIV-1 prompts a massive cellular immune response, but eventually it tires the immune cells. Blocking the activation of a cell receptor called PD-1 might restore these exhausted cells.

“There are only the pursued, the pursuing, the busy, and the tired.”

F. Scott Fitzgerald *The Great Gatsby*.

When confronted with a pathogen, the immune system must respond with appropriate force, sufficient to clear the pathogen but not so vigorous as to damage the host through excessive inflammation. So, the effector arms of the immune system — the cells and molecules that carry out the immune response — are carefully regulated to achieve a balance between pathogen clearance and immunopathology. Failure to achieve this balance can lead to persistent infection, often accompanied by chronic inflammation and disease, as seen in human infection with hepatitis viruses B and C and the human immunodeficiency virus HIV-1. Moreover, the mechanisms that control the balance between effective response and immune-mediated damage to the host can be subverted by pathogens — through millennia of coevolution, pathogens have acquired substantially more profound insights into the complexity of host immune responses than human immunologists have yet achieved.

Maintaining the balance of the immune response is especially critical in regulating T cells, which secrete inflammatory molecules (cytokines) and kill infected cells. In chronic infection, the responding T cells often show

features of what immunologists term ‘exhaustion’, with varying degrees of impairment of their effector functions. A paper in this issue (page 350)¹ by Day *et al.* and one by Trautman *et al.* just published on the *Nature Medicine* website² show how an essential immune mechanism that regulates the response of T cells is undermined in HIV-1 infection.

One of the unexplained paradoxes in HIV infection has been why a huge cellular immune response — in which as many as one in five circulating killer T cells are responding to HIV antigens — fails to control or clear the virus. Nevertheless, most immunologists believe that virus-specific killer T cells are a key component of immunity to HIV-1, particularly in the early stages of infection, and many of the HIV vaccine candidates currently in the early stages of clinical trials have been designed to stimulate a killer-T-cell response. Can the paradox be explained by a progressive loss of function of HIV-specific killer T cells (as suggested in some studies)^{3–5} — and, more importantly, can this process be reversed?

The mouse model of chronic infection with lymphocytic choriomeningitis virus (LCMV) has many parallels with human HIV-1 infection. Earlier this year, Barber *et al.*⁶ used this model to show that exhausted killer T cells are characterized by expression of an inhibitory molecule called programmed death-1 (PD-1;

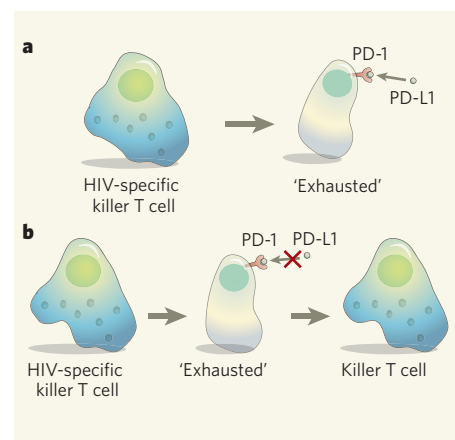


Figure 1 | Immune exhaustion and PD-1.

a, Although immune cells called T cells initially respond to an HIV infection by killing the infected cells and secreting inflammatory molecules, they seem to gradually give up the fight against the virus. In a mouse model of the disease, these ‘exhausted’ T cells are characterized by expression of the PD-1 receptor on their cell surface. **b**, Day *et al.*¹ and Trautman *et al.*² studied the role of PD-1 in cultured cells from HIV-1-infected patients. Day *et al.* show that blocking the interaction between PD-1 and its ligand (PD-L1) seems to restore certain of the T-cell responses, at least at the level of the bulk culture.

Fig. 1a). This cell-surface molecule is part of the CD28 family and is thought to be a regulator of the T-cell response to invading pathogens, interacting with its ligands (PD-L1 and PD-L2) to inhibit activation of T cells once the T-cell surface receptors have recognized the virus.

This mechanism probably has a physiological role in limiting the expansion in the numbers of T cells responding to an acute infection. But in the dysregulated setting of chronic infection, PD-1 expression on virus-specific T cells

seems to be a signal that the T cells are giving up the fight and can no longer perform vital functions to control the infection. Remarkably, in the LCMV-infected mice, administration of an antibody that blocked the interaction of PD-1 with its ligands allowed these exhausted T cells to regain the ability to secrete cytokines, to kill other cells and to proliferate, leading to control of the virus. Moreover, this strategy was effective even in animals lacking 'T-cell help', the essential supportive functions provided by T cells expressing the CD4 molecule that makes them a prime target for HIV infection. This raises the possibility that such an antibody regime might be relevant in addressing HIV-1 infection, as a cardinal feature of the infection is a lack of helper T cells.

Following on from this important finding, Day *et al.*¹ studied a cohort of untreated HIV-1-infected patients in South Africa. They now report that PD-1 is significantly upregulated on killer T cells in the blood that recognize HIV-1 antigens. Both the proportion of killer T cells expressing PD-1 and the levels of PD-1 on the cell surface showed a strong correlation with the viral load in the blood plasma — currently the best indicator of disease progression — and decreased when the viral load was controlled by anti-retroviral therapy. More strikingly, they found that by blocking the interaction between PD-1 and its ligands, several effector functions of killer T cells, notably proliferation and the ability to secrete antiviral cytokines, could be restored *in vitro* (at least in bulk cell culture, if not on an individual cell basis; Fig. 1b).

Similar studies by Trautmann and colleagues² show restoration of a range of effector molecules in the presence of an antibody that blocks the PD-1 ligands. These experiments confirm that functional exhaustion is a feature of virus-specific T cells that are responding to HIV-1 infection, and show that impaired T-cell function is magnified by increasing exposure to replicating virus.

Both reports^{1,2} show that PD-1 expression levels vary in different T-cell populations, even within the same individual. This raises the possibility that some responding T cells can maintain function better than others, as described in long-term HIV-infected patients who do not develop disease or any signs of immune suppression⁷.

The exciting conclusion from both studies is that, despite the damage to T cells that results from their encounter with HIV-1, the cells are not terminally incapacitated and can potentially be revived. The work will undoubtedly prompt therapeutic strategies to restore responding T cells to full health and function, not only in HIV-1 infection, but in other settings where T-cell exhaustion is believed to play a part in disease, such as chronic infection with hepatitis B and C viruses. Similar approaches are being explored in cancer immunotherapy⁸, because upregulation of the PD-1 ligands is a feature of many tumours and is associated with a poor prognosis.

These strategies will need to be approached with caution, as studies in mice suggest that blocking the interactions between PD-1 and its ligands can exacerbate autoimmune disease. Intriguingly, several human autoimmune conditions are associated with a mutation (a single nucleotide polymorphism) in the regulatory region of the *PD-1* gene⁹. T cells from donors with this genetic variant seem to be less susceptible to PD-1-mediated suppression of T-cell function⁹, so could this mutation provide an advantage to a host infected with a persistent virus or faced with a tumour that overexpresses a PD-1 ligand? Just as PD-1 ligand expression seems to be a manoeuvre on the part of tumours to avoid recognition by T cells, several viruses have developed mechanisms to upregulate PD-1 ligand in the tissues. HIV-1 is already known to possess an extraordinarily diverse

repertoire of immune evasion strategies: will the stimulation of PD-1 ligand expression in T cells prove to be a further example of viral subterfuge?

Sarah Rowland-Jones and Tao Dong are in the MRC Human Immunology Unit, Weatherall Institute of Molecular Medicine, Oxford OX3 9DS, UK. Sarah Rowland-Jones is also at the MRC Laboratories, Fajara, Gambia.

e-mail: sarah.rowland-jones@ndm.ox.ac.uk

1. Day, C. L. *et al.* *Nature* **443**, 350–354 (2006).
2. Trautmann, L. *et al.* *Nature Med.* doi:10.1038/nm1482 (2006).
3. Appay, V. *et al.* *J. Exp. Med.* **192**, 63–75 (2000).
4. Shankar, P. *et al.* *Blood* **96**, 3094–3101 (2000).
5. Kostense, S. *et al.* *Eur. J. Immunol.* **31**, 677–686 (2001).
6. Barber, D. L. *et al.* *Nature* **439**, 682–687 (2006).
7. Migueles, S. A. *et al.* *Nature Immunol.* **3**, 1061–1068 (2002).
8. Geng, H. *et al.* *Int. J. Cancer* **118**, 2657–2664 (2006).
9. Kroner, A. *et al.* *Ann. Neurol.* **58**, 50–57 (2005).

ASTRONOMY

Champagne supernova

David Branch

Thermonuclear supernovae were thought to occur only when white-dwarf stars of a certain mass explode. The discovery of a supernova that is way over the mass limit might require a reworking of the model.

According to the conventional view, thermonuclear, or type Ia, supernova explosions occur when a white-dwarf star — a star that has exhausted its nuclear fuel and is composed entirely of carbon and oxygen — accretes matter from a close companion star. At the same time, the white dwarf contracts, and so its density and temperature increase. This goes on until its mass approaches the maximum that a white dwarf can have — the Chandrasekhar mass, 1.4 times that of the Sun. At this point, a violent thermonuclear instability releases enough energy from nuclear fusion to blast the white dwarf apart at speeds of a few per cent of the speed of light. The nuclear fusion burns about 0.6 solar masses of the white dwarf to a single isotope, radioactive nickel-56. The decay of this isotope, through cobalt-56 to stable iron-56, provides a delayed source of energy that keeps the ejected matter hot, causing the supernova to attain a peak luminosity (defined as energy radiated per unit time) greater than that of a billion Suns.

On page 308 of this issue, Howell *et al.*¹ throw a spanner in the works. They report observations of a supernova that was too luminous to have been produced by a mere 1.4 solar masses of ejecta. This is the first conclusive evidence for a 'super-Chandra' supernova.

Type Ia supernovae are of great interest to astronomers because they are used to investigate the expansion history of the Universe. Their high luminosities mean they can be seen far away (and therefore far back in time), and

their relative brightness can be used to infer their distance. An empirical relation between the rate at which the supernova's luminosity rises and falls, known as the 'stretch'², and its peak luminosity allows the latter value, and so the distance, to be inferred with great precision.

In the past decade, type Ia supernovae have been used to show that the expansion of the Universe must be accelerating. This process, working against the pull of gravity, must be driven by some mysterious 'dark energy', possibly supplied by the phenomenon represented by Einstein's cosmological constant. Astronomers are planning to observe large numbers of type Ia supernovae to study the nature of this dark energy^{3–6}, and are thus eager to know more about the physical features of these titanic nuclear-powered stellar explosions.

The supernova discovered by Howell *et al.*¹ — called SNLS-03D3bb, and also known as SN 2003fg — has a spectrum of emissions and absorptions at optical wavelengths that establishes it as being of type Ia. This means that it is powered by a runaway thermonuclear reaction as described above, rather than by the gravitational processes that power other supernova types. But the peak luminosity of the supernova was 2.2 times higher than that of a typical type Ia event. This luminosity depends on the mass of nickel-56 present; in this case, about 1.3 solar masses of nickel would be needed to produce such luminosity.

That amount of nickel requires far more than

the 1.4 solar masses of initial ejecta allowed by the Chandrasekhar limit. This is because fusion reactions produce not just nickel, but also stable iron-group isotopes, and the optical spectrum of SN 2003fg reveals the presence of lighter elements such as silicon, sulphur and calcium. There may also be unburnt oxygen and carbon. Taking all this together, Howell *et al.* estimate that the mass of the ejecta must have been 2.1 solar masses.

How can a white dwarf be so massive? One way would be if two white dwarfs were to spiral together and eventually coalesce. Such an event could occur through loss of a star's angular momentum, caused by ripples in space-time known as gravitational waves. But such an object would probably collapse to a neutron star^{7,8} rather than explode as a supernova. A more likely explanation is that matter accumulated by a white dwarf from a normal companion star adds extra angular momentum, causing the white dwarf to rotate more rapidly. This rapid rotation would provide additional support against gravity and allow the white dwarf to become overmassive before it exploded.

The maximum mass a white dwarf might achieve in this way depends on how the angular momentum is distributed within the star; that is, whether it rotates as a single, rigid body, or whether different parts of it rotate at different rates. Differential rotation might be able to support as much as four solar masses, although limitations on how much mass the companion star can transfer may constrain white-dwarf masses to not much more than two solar masses⁹.

As a result of the Doppler effect — in which the radiation emitted by an object moving away is stretched or 'redshifted' — features in the spectrum of SN 2003fg are broadened, but by an amount that shows that the ejection velocity was lower than is typical of type Ia supernovae¹. That is consistent with a super-Chandra white dwarf, because although more nuclear energy is released by fusion in such a body, a higher binding energy is also required to break it up against its self-gravity. Higher binding energy can lead to lower ejection velocity.

To further our understanding of type Ia supernovae, it will be crucial to determine the distribution of masses that they eject. For example, are type Ia supernovae generally super-Chandra, with a smooth distribution between 1.4 and 2.1 solar masses? Howell *et al.*¹ present an analysis of a supernova sample hinting that this may be so. The present data, however, are compatible with the typical type Ia supernova being near the Chandra mass, with SN 2003fg as a special case. This interpretation is suggested by the supernova's distinct overluminosity compared with its peers, and by its violation of the luminosity–stretch relation: despite its extreme brightness, its stretch is typical.

That need not mean that SN 2003fg raises a problem for the use of type Ia supernovae

as cosmic distance indicators. If one assumed that SN 2003fg conformed to the luminosity–stretch relation, one would severely underestimate its luminosity, and thus its distance. But it is such a blatant outlier that it has already been excluded from consideration in one cosmology study¹⁰. The luminosity–stretch relation is empirical and does not entail assumptions about the mass distribution of type Ia supernovae. So it may already be able to accommodate some deviations from the Chandra mass — although not for the exceptional SN 2003fg.

David Branch is in the Homer L. Dodge Department of Physics and Astronomy,

University of Oklahoma, Norman, Oklahoma 73019, USA.
e-mail: branch@nhn.ou.edu

1. Howell, D. A. *et al.* *Nature* **443**, 308–311 (2006).
2. Goldhaber, G. *et al.* *Astrophys. J.* **558**, 359–368 (2001).
3. Wood-Vasey, W. M. *et al.* *N. Astron. Rev.* **48**, 637–640 (2004).
4. Corasaniti, P. S., Lo Verde, M., Crotts, A. & Blake, C. *Mon. Not. R. Astron. Soc.* **369**, 798–804 (2006).
5. Tyson, J. in *Observing Dark Energy* (eds Wolff, S. C. & Lauer, T.) 95 (Astron. Soc. Pacif., San Francisco, 2005).
6. Aldering, G. N. *Astron. Rev.* **49**, 346–353 (2005).
7. Saio, H. & Nomoto, K. *Astrophys. J.* **615**, 444–449 (2004).
8. Piersanti, L., Gagliardi, S., Iben, I. Jr & Tornambé, A. *Astrophys. J.* **598**, 1229–1238 (2003).
9. Yoon, S.-C. & Langer, N. *Astron. Astrophys.* **435**, 967–985 (2005).
10. Astier, P. *et al.* *Astron. Astrophys.* **447**, 31–48 (2006).

STEM CELLS

A new route to rejuvenation

M. Azim Surani and Anne McLaren

Embryonic stem cells are prized for their ability to mature into all the specialized adult cell types. It may now be possible to reprogramme adult body cells to have the characteristics of stem cells.

It is relatively easy to grow an entire plant from a small cutting, but it seems inconceivable that the diverse tissues such as nerve and muscle in mammals like ourselves could be regenerated directly from a single adult tissue. Yet a study by Takahashi and Yamanaka published in *Cell*¹ has brought us tantalizingly close to just such a possibility. Working with a variety of fetal and adult mouse cells, including adult fibroblasts from tail clippings, these investigators rejuvenated them using just four genes, restoring their vigour and their potential to differentiate into diverse cell types. This remarkable work

improves the prospects for a similar outcome using human cells.

Embryonic stem cells (ES cells), derived from 4-day-old mouse or human embryos (blastocysts), have the exceptional ability to either multiply indefinitely (self-renewal) or differentiate into all the adult cell types, depending on their culture conditions. As embryos develop further, their cells become progressively specialized, and it eventually becomes impossible to generate ES cells from the fetus. Specialized cells, such as those of nerves and muscle, become different from each other through the expression of a 'network' of genes unique

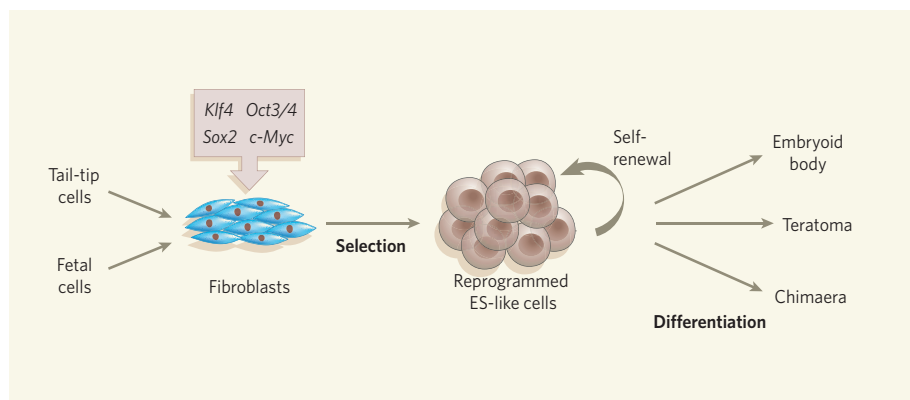


Figure 1 | Cellular reprogramming. Takahashi and Yamanaka¹ show that specialized cells from fetal or adult mice can be rejuvenated and reprogrammed into cells resembling embryonic stem (ES) cells. The authors took defined mouse cells (fibroblasts) from fetuses or adult tail-tips and cultured them. They then introduced four key genes, *Klf4*, *Oct3/4*, *Sox2* and *c-Myc*, into the cells. This is followed by a selection process to isolate the small fraction of reprogrammed cells. These cells resemble ES cells in that they display the key properties of self-renewal and the ability to mature into many different cell types. The latter were observed in embryoid bodies (cell aggregates formed by the reprogrammed cells in culture), in teratomas (tumours generated from the reprogrammed cells), and in chimaeric fetuses made by implanting reprogrammed cells into a host embryo. Some of the chimaeras developed beyond the mid-gestation period, but no adults were obtained.

to the cell type. The cells' individual identity is robustly maintained thereafter by heritable (but potentially reversible) chemical modifications both to the genes themselves and to proteins that are bound to the DNA. These modifications are referred to as 'epigenetic' because they regulate the all-important genetic networks without altering the actual genetic constitution of the cells.

So how did Takahashi and Yamanaka break through these formidable barriers that maintain the differentiated state of specialized cells and convert the cells into reprogrammed ES-like cells? Essentially, their approach was to initiate a new transcriptional network. To do this, they isolated 24 genes that are apparently necessary for the unique properties of ES cells, and introduced extra copies of all these genes together into cultured fetal and adult mouse cells. A very small fraction of cells showed ES-like characteristics. Remarkably, the authors eventually found that only four genes — *Oct3/4*, *Sox2*, *c-Myc* and *Klf4* — are sufficient for this conversion (Fig. 1). The first two are well known as genes that encode gene-regulatory proteins that are necessary to maintain ES-like characteristics^{2,3}. The *c-Myc* and *Klf4* proteins support the self-renewal of ES cells⁴, and *Klf4* also augments the levels of *Oct4* (ref. 5).

One key gene-regulatory protein — Nanog — is notably absent from the list. Nanog can enhance reprogramming⁶, and, together with *Oct3/4* and *Sox2*, constitutes the core set of gene-regulatory factors in ES cells^{3,4}. Nanog was detected in most of the reprogrammed cells, however, and so the introduced *Oct3/4* and *Sox2* may have stimulated its activity in these cells⁷. Indeed, the combined effects of the four genes identified must induce extensive changes in the overall transcriptional network in the cells.

A detailed 'fingerprint' of the gene expression pattern in reprogrammed cells confirms similarities between them and the ES cells, but significant deviations are also evident¹. Nevertheless, the reprogrammed cells display the crucial capacity for self-renewal and differentiation into many types of tissue that is characteristic of ES cells (Fig. 1). Furthermore, when introduced into a host embryo, the reprogrammed cells in the resulting chimaeric fetuses differentiate into many types of tissue. Unlike the case with ES cells, however, none of the chimaeras developed to adulthood, which suggests imperfections in the reprogramming. Notably, the target cells' own copies of the *Oct3/4* and *Sox2* genes were not significantly expressed after the reprogramming, and the cells instead relied on the activity of the newly introduced genes to maintain their ES-cell properties. Moreover, the cells' own *Oct3/4* gene showed significant DNA methylation¹, which is an epigenetic hallmark of an inactive gene and unlike the situation in ES cells. The epigenetic modifications in the reprogrammed cells thus seem to differ somewhat from

those in ES cells, which have their own unique epigenetic characteristics essential for their properties and plasticity^{8,9}.

So what of the future? It seems that there are at least two major requirements for efficient reprogramming of adult cells. One of these is clearly to kick-start a new transcriptional network through the introduction of crucial factors such as *Oct3/4* and *Sox2*. But before this, erasing at least some of the existing epigenetic modifications from the specialized cells¹⁰ might make it easier to set up the required changes in the transcriptional network. If this erasure could be achieved, it might make the procedure more efficient, such that the introduced genes might be needed only transiently to achieve full reprogramming of adult cells.

In principle, the technique devised by Takahashi and Yamanaka¹ should work with human cells because the transcriptional factors needed to maintain human ES cells are apparently similar to those in mice^{3,4}. The environmental factors and cues needed for the derivation of mouse and human ES cells are different, however, and their role during cell reprogramming has not been fully worked out. But if successful, the procedure could eventually be used to generate reprogrammed stem-like cells from adult human body cells for use in therapy. Reprogrammed cells from patients with complex diseases, such

as diabetes, would also be a valuable resource for research. They could be used to generate differentiated tissues to see how different mutations and genetic backgrounds influence the progression of the disease — information that might lead to prevention or a cure. Before this can happen, however, we must address the gaps in our knowledge, so that we can derive fully reprogrammed cells efficiently. This approach may eventually eliminate the need to use early embryos for deriving stem cells — an enticing objective, but one that will require extensive research on both mouse and human ES cells. ■

M. Azim Surani and Anne McLaren are at the Wellcome Trust Cancer Research UK Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK. e-mails: as10021@mole.bio.cam.ac.uk; a.mclaren@gurdon.cam.ac.uk

1. Takahashi, K. & Yamanaka, S. *Cell* **126**, 663–676 (2006).
2. Ilvanova, N. *et al.* *Nature* **442**, 533–538 (2006).
3. Boyer, L. A. *et al.* *Cell* **122**, 947–956 (2005).
4. Cartwright, P. *et al.* *Development* **132**, 885–896 (2005).
5. Li, Y. *et al.* *Blood* **105**, 635–637 (2005).
6. Silva, J., Chambers, I., Pollard, S. & Smith, A. *Nature* **441**, 997–1001 (2006).
7. Kuroda, T. *et al.* *Mol. Cell. Biol.* **25**, 2475–2485 (2005).
8. Azura, V. *et al.* *Nature Cell Biol.* **8**, 532–538 (2006).
9. Bernstein, B. E. *et al.* *Cell* **125**, 315–326 (2006).
10. Surani, M. A. *Cell* **122**, 653–654 (2005).

CHEMICAL BIOLOGY

A sweet exchange

Christopher T. Walsh

Many drugs isolated from microorganisms have complex molecular structures, making it difficult for chemists to modify them. But it seems that enzymes can provide a short cut to drug variants.

Naturally occurring compounds — referred to as natural products by chemists — are excellent starting points for making new medicines. Many drugs hail from natural products isolated from microorganisms: for example, the erythromycin and vancomycin classes of antibiotic, the daunomycin and adriamycin classes of anticancer drug, and the so-called 'enediynes warheads' that have been attached to therapeutic antibodies, which are also used in anticancer therapy.

Nature makes these molecules biologically active by building in conformational constraints, which set specific architectures into the molecular 'scaffolds'. The peripheries of the scaffolds are then decorated with chemical groups that are recognized by biological receptors. Dedicated tailoring enzymes modify the nascent natural-product scaffolds at distinct locations, bringing in methyl groups, acyl groups and a range of unusual sugars that become key recognition elements for biological targets. Chemists dearly want to introduce

these rare sugars into natural-product scaffolds at will, varying the point of attachment in the hope of making improved drugs, but making these sugars is extremely difficult. This problem may now be solved, as Zhang *et al.* report in *Science*¹ that sugars can be enzymatically harvested from natural products and readily swapped between scaffolds.

The sugars attached to natural products serve many functions. Because they contain hydroxyl (OH) groups, sugars are hydrophilic, so their incorporation into a natural product can make the overall molecule water-soluble. The arrays of hydroxyl groups might also act as hydrogen-bonding elements for specific interactions with biological macromolecules².

The appended sugars can be widespread in the natural world — for example glucose and mannose, which belong to a family of sugars known as D-hexoses. But most often they are specialized, such as L-vancosamine in vancomycin, desosamine in erythromycin or L-daunosamine in daunomycin. These sugars

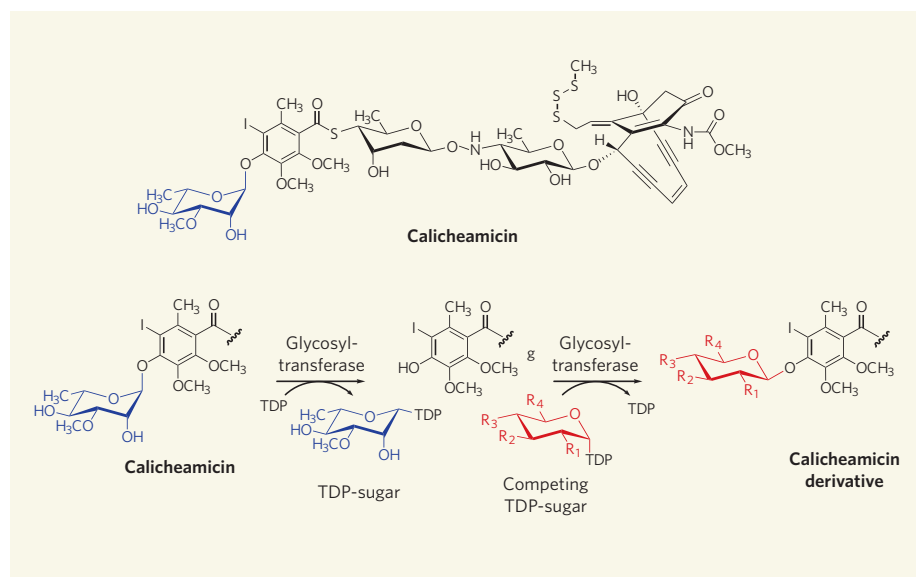


Figure 1 | Enzymatic swap of sugars between antibiotic scaffolds. As reported by Zhang *et al.*¹, a glycosyltransferase (GT) enzyme catalyses the removal of a sugar (blue) from a molecule of the anticancer drug calicheamicin, in the presence of deoxythymidine diphosphate (TDP), producing a TDP-sugar as a side-product; R_1 – R_4 represent general chemical groups. If this reaction is carried out in the presence of a competing TDP-sugar, the GT will attach that sugar (red) to the drug scaffold, effecting a net sugar-exchange reaction. In this way, variants of calicheamicin are produced that would be extremely difficult to make by chemical synthesis.

differ from D-hexoses by lacking a particular hydroxyl group (deoxyhexoses) or by lacking a hydroxyl and gaining an amine (NH_2) group (aminodeoxyhexoses). The variations in chemical structure create a variable hydrophilic or hydrophobic balance in the resulting sugar, providing a unique chemical code that can be read by target proteins.

The enzymes that synthesize these specialized sugars are encoded by DNA that lies close to the gene clusters that encode the scaffold-producing enzymes. This proximity allows sugar synthesis to be regulated so that the sugars are produced only when the scaffolds are ready for them. D-Hexoses provide the raw material from which the rare sugars are prepared. The hexoses are converted into reactive intermediates, known as deoxythymidine diphosphohexoses (TDP-sugars), which are then acted on by dedicated glycosyltransferase (GT) enzymes. The GTs transfer sugars in the form of TDP-hexoses to specific sites in the natural-product scaffold, so creating an active final product. For example, in the biogenesis of vancomycin, six enzymes are needed to convert TDP-glucose to TDP-vancosamine, whereupon two GTs attach the sugar to the vancomycin scaffold. Frustratingly for drug developers, vancosamine, like desosamine in erythromycin and daunosamine in daunomycin, is not readily available by synthetic routes^{3,4}.

It had long been assumed that GT-catalysed reactions are irreversible: TDP-sugars react with a scaffold to yield an end-product, with TDP produced as a side-product. But Zhang *et al.*¹ have serendipitously discovered reversible reactions with a GT known as CalG1. While using this enzyme to modify the core

of the antitumour agent calicheamicin, they found that an appended sugar within the drug could be transferred back to TDP. When they attempted this transfer in the presence of a particular TDP-deoxyglucose, the sugar in calicheamicin was replaced with that deoxyglucose: a net sugar-exchange reaction had occurred (Fig. 1). This is an unexpected and extremely useful observation.

Zhang *et al.*¹ found that a variety of hexoses may be installed on the calicheamicin scaffold using this GT-mediated sugar exchange. They then extended the process by harvesting exotic sugars from other natural products and appending them to modified calicheamicin scaffolds in one reaction. In this way, they prepared a library of 70 calicheamicin analogues. Creating a library of this size by traditional chemical synthesis would be extremely time-consuming. Even more impressively, Zhang *et al.*¹ identified a second enzyme, CalG4, which acts at a different part of the calicheamicin scaffold; CalG4 removes a different sugar (an aminopentose) from this position in the scaffold, and also catalyses sugar-exchange reactions. Combinations of sugar-exchange reactions using CalG1 and CalG4 could presumably give libraries with high structural diversity, by mixing pairs of hexoses and aminopentoses on the calicheamicin core.

Perhaps most encouragingly, the authors¹ expanded the scope of the sugar-exchange reactions by using GTs associated with other natural products. The enzyme GtfE decorates the scaffold of the antibiotic vancomycin with D-glucose. Incubations of GtfE and CalG1 together in the presence of TDP allow the first enzyme to remove a modified glucose

(azidoglucose) from a variant of vancomycin, transiently generating TDP-azidoglucose. This short-lived intermediate is then used by CalG1 to transfer azidoglucose to a calicheamicin scaffold with a good overall yield (albeit on a very small reaction scale of 5 nmol of calicheamicin). This constitutes a simple enzymatic swap of sugars between completely different natural products. It is early days, but if these reactions can be run on a larger scale, GT-mediated sugar exchange might provide a short cut for shuffling biologically active sugars around natural-product scaffolds. This could be useful for finding drug molecules for many therapeutic areas.

Evaluating the generality and utility of this approach still presents several challenges. The first is to test a larger set of GTs as catalysts for sugar-exchange reactions. More than a hundred microbial GTs are noted in genome databases. If many of these enzymes can indeed be used in this way, a door opens to a mix-and-match strategy for creating new sugar forms of therapeutic natural products.

A second challenge is to obtain a ready supply of the rare deoxyhexose and aminodeoxyhexose sugars that are required for running the reactions. These sugars can only be obtained from their parent natural products, which are invariably generated from large-scale fermentations of producer microorganisms. Most of these natural products are not routinely available, but if Zhang and colleagues' method is widely adopted, this might put a premium on fermentation scale-ups.

Many enzymes exist that decorate natural products with chemical units such as methyl groups, acyl groups and occasionally phosphoryl groups⁵. These enzymes should also be examined for the purpose of transferring chemical groups between natural products. For example, the net transfer of acyl groups between natural-product scaffolds by acyltransferase enzymes would, in combination with sugar swaps, produce building-blocks for libraries of molecules with potential therapeutic activities.

By following up a chance observation of unexpected enzyme activity, Zhang *et al.* have developed a tool for selectively modifying exceedingly complex molecules. If this is a truly general process, then the long-held dreams of many carbohydrate chemists have been surpassed.

Christopher T. Walsh is in the Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA.
e-mail: christopher_walsh@hms.harvard.edu

1. Zhang, C. *et al.* *Science* **313**, 1291–1294 (2006).
2. Walsh, C. T., Losey, H. C. & Freil Meyers, C. L. *Biochem. Soc. Trans.* **31**, 487–492 (2003).
3. Oberthur, M., Leimkuhler, C. & Kahne, D. *Org. Lett.* **6**, 2873–2876 (2004).
4. Oberthur, M. *et al.* *J. Am. Chem. Soc.* **127**, 10747–10752 (2005).
5. Fischbach, M. A. & Walsh, C. T. *Chem. Rev.* **106**, 3468–3496 (2006).

BRIEF COMMUNICATIONS

Induction of an illusory shadow person

Stimulation of a site on the brain's left hemisphere prompts the creepy feeling that somebody is close by.

The strange sensation that somebody is nearby when no one is actually present has been described by psychiatric and neurological patients, as well as by healthy subjects, but it is not understood how the illusion is triggered by the brain^{1,2}. Here we describe the repeated induction of this sensation in a patient who was undergoing presurgical evaluation for epilepsy treatment, as a result of focal electrical stimulation of the left temporoparietal junction: the illusory person closely 'shadowed' changes in the patient's body position and posture. These perceptions may have been due to a disturbance in the multisensory processing of body and self at the temporoparietal junction.

The patient was a 22-year-old woman of normal psychiatric history who was undergoing evaluation for surgical treatment of epilepsy (see supplementary information). We identified an area on the left temporoparietal junction in her brain (Fig. 1a) where focal electrical stimulation repeatedly produced a feeling of the presence of another person in her extrapersonal space.

When stimulated at this region (for methods, see supplementary information; current amplitude, 10.0 mA) in a supine position, the patient had the impression that somebody was behind her. Further stimulation (10.0–11.0 mA; $n = 3$) induced the same experience, with the patient describing the "person" as young and of indeterminate sex, a "shadow" who did not speak or move, and whose position beneath her back was identical to her own ("He is behind me, almost at my body, but I do not feel it"; Fig. 1b).

During the next stimulation (11.0 mA; $n = 1$), the patient sat and embraced her knees with her arms (Fig. 1c). She noted that the "man" was now also sitting and that he was clasping her in his arms, which she described as an unpleasant feeling. Further stimulations (11.0 mA; $n = 2$) were applied while the seated patient performed a naming (language-testing) task using a card held in her right hand (Fig. 1d): she again reported the presence of the sitting "person", this time displaced behind her to her right and attempting to interfere with the execution of her task ("He wants to take the card"; "He doesn't want me to read"). Similar effects were observed for different positions and postures (see supplementary information) when stimuli exceeding 10 mA were applied to the same site on the left temporoparietal junction.

The sensation of a presence, as reported by this patient, has been described in people

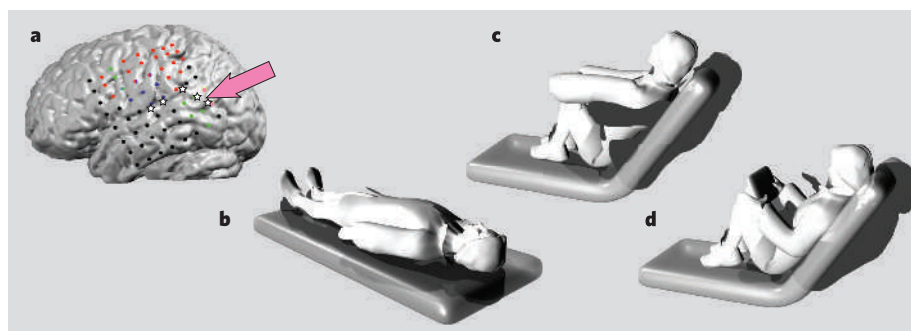


Figure 1 | Me and my shadow. **a**, Three-dimensional surface reconstruction of the left hemisphere of the brain from magnetic resonance imaging. Subdural electrodes were implanted in the brain of an epileptic patient undergoing presurgical evaluation¹⁰. The locations at which focal electrical stimulation evoked different responses are shown (red, motor; blue, somatosensory; green, language; pink, site where 'feeling of a presence' could be induced (arrow)); stars indicate the epileptic focus. Since undergoing a left-temporal lobectomy, the patient has been free of seizures. **b–d**, Drawings showing relative positions and postures of the patient's body (white) and that of the illusory person (shaded) during cortical stimulation (illustrations, M. Boyer).

with psychiatric and neurological disorders^{1–4}. Because it was possible to induce this feeling repeatedly from the temporoparietal junction, and because the illusory person closely mimicked the patient's body posture and position, we conclude that the patient was experiencing a perception of her own body^{1,4}.

The temporoparietal junction is known to be involved in self-processing, self-other distinction, the integration of multisensory body-related information, and other illusory own-body perceptions^{5,6}. We therefore propose that electrical stimulation of this area in our patient disturbed multisensory (proprioceptive and tactile) and sensorimotor integration of information with respect to her body, leading to the appearance of a first-rank schneiderian symptom of schizophrenia in a person with no psychiatric history. It is notable that hyperactivity in the temporoparietal cortex of patients with schizophrenia may lead to the misattribution of their own actions to other people⁷.

In conclusion, we argue that multisensory and/or sensorimotor disintegration⁸ at the temporoparietal junction due to electrical stimulation may lead to an own-body illusion of another person in near extrapersonal space. Although our patient was aware of the similarity between her own postural and positional features and those of the illusory person, she did not recognize that that person was an illusion of her own body, like many deluded schizophrenic patients^{3,8,9}. Our findings may be a step towards understanding the mechanisms behind psychiatric manifestations such as

paranoia, persecution and alien control. They also highlight the importance of multimodal mechanisms at the temporoparietal junction in human self-attribution.

Shahar Arzy*‡, Margitta Seeck†‡, Stephanie Ortigue§, Laurent Spinelli†, Olaf Blanke*‡

*Laboratory of Cognitive Neuroscience, École Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland
e-mail: olaf.blanke@epfl.ch

†Presurgical Epilepsy Evaluation Unit,

‡Department of Neurology, University Hospital, Geneva 1211, Switzerland

§Center for Cognitive Neuroscience, Dartmouth College, Dartmouth, New Hampshire 03755, USA

1. Brugger, P., Regard, M. & Landis, T. *Neuropsychiatr. Neuropsychol. Behav. Neurol.* **9**, 114–122 (1996).
2. Critchley, M. *Acta Psychiatr. Neurol. Scand.* **30**, 155–168 (1955).
3. Jaspers, K. *General Psychopathology* (Manchester Univ. Press, Manchester, 1972).
4. Blanke, O., Ortigue, S., Coeytaux, A., Martory, M. D. & Landis, T. *Neurocase* **9**, 329–339 (2003).
5. Blanke, O., Landis, T., Spinelli, L. & Seeck, M. *Brain* **127**, 243–258 (2004).
6. Ruby, P. & Decety, J. *Nature Neurosci.* **4**, 546–550 (2001).
7. Farrer, C. et al. *Psychiatr. Res.* **131**, 31–44 (2004).
8. Blakemore, S.-J., Wolpert, D. M. & Frith, C. D. *Trends Cogn. Sci.* **6**, 237–242 (2002).
9. Thompson, C. B. *J. Psychiatr.* **141**, 628–630 (1982).
10. Blanke, O., Perrig, S., Thut, G., Landis, T. & Seeck, M. *J. Neurol. Neurosurg. Psychiatr.* **69**, 553–556 (2000).

Supplementary information accompanies this communication on Nature's website.

Received 19 July; accepted 25 August 2006.

Competing financial interests: declared none.
doi:10.1038/443287a

BRIEF COMMUNICATIONS ARISING online
www.nature.com/bca see Nature contents.

GENE THERAPY

Is *IL2RG* oncogenic in T-cell development?Arising from: N.-B. Woods, V. Bottero, M. Schmidt, C. von Kalle & I. M. Verma *Nature* **440**, 1123 (2006)

The gene *IL2RG* encodes the γ -chain of the interleukin-2 receptor and is mutated in patients with X-linked severe combined immune deficiency (X-SCID). Woods *et al.*¹ report the development of thymus tumours in a mouse model of X-SCID after correction by lentiviral overexpression of *IL2RG* and claim that these were caused by *IL2RG* itself. Here we find that retroviral overexpression of *IL2RG* in human CD34⁺ cells has no effect on T-cell development, whereas overexpression of the T-cell acute lymphoblastic leukaemia (T-ALL) oncogene *LMO2* leads to severe abnormalities. Retroviral expression of *IL2RG* may therefore not be directly oncogenic — rather, the restoration of normal signalling by the interleukin-7 receptor to X-SCID precursor cells allows progression of T-cell development to stages that are permissive for the pro-leukaemic effects of ectopic *LMO2*.

The occurrence of leukaemia in three patients in a gene-therapy trial for X-SCID (refs 2–4) has highlighted an adverse effect of insertional mutagenesis as a delivery technique for the therapeutic gene. In our investigation, we expressed both *LMO2* and *IL2RG* transgenes in the same retroviral vector that was used in the clinical gene-therapy trials; those trials found that retroviral integration of the corrective *IL2RG* occurred near the locus of the *LMO2* oncogene⁵. We concluded that this integration may have upregulated the expression of *LMO2* and triggered the leukaemia cases in the gene-therapy trials.

Woods *et al.* used a murine model in which stem cells from wild-type or mutant *Il2rg*^{-/-} mice were transduced with very high amounts of human *IL2RG* by means of a lentiviral vector¹. They found that these transplanted mice developed a high incidence of tumours. To the best of our knowledge, however, *IL2RG* has never been reported to act as an oncogene in human T-ALL, and murine studies should be interpreted with caution when extrapolating to humans. Although T-cell development is very similar in mice and humans at the genome-wide level^{6,7}, there are differences in individual genes: for instance, mutant *Il2rg*^{-/-} mice lack T, B and natural killer (NK) cells, whereas human X-SCID patients have normal numbers of B cells.

Mouse–human hybrid fetal thymus organ culture (FTOC) is often used to study human T-cell development *ex vivo*. In FTOC experiments, we found that *LMO2*-transduced cells expressed the early thymocyte marker CD1a less frequently than untransduced cells derived from the same culture (22% compared with 57%), whereas a higher proportion of *LMO2*-transduced cells retained the progenitor marker CD34 (12% compared with 2–4%, representa-

tive of five different FTOC experiments). Later, during T-cell development, this block was even more pronounced (58% of cells in the subsequent immature CD4 single-positive stage, compared with 22% in controls). These results indicate that progenitor cells ectopically expressing *LMO2* are severely hampered in T-cell development.

The percentage of immature CD4 single-positive (37% compared with 38%) and the later CD4⁺CD8⁺ double-positive cells (6% compared with 9%) was similar in untransduced and *IL2RG*-transduced populations, indicating that retroviral-mediated expression of *IL2RG* does not affect T-cell development. *LMO2*-transduced cells developed normally into B, NK and myeloid cells.

The very high expression of lentiviral transgenic *IL2RG* in the transplanted mice studied by Woods *et al.*¹ may have contributed to development of T-cell lymphomas. In contrast, *IL2RG* concentrations were slightly lower than normal in the human X-SCID trials after treatment using retroviral vectors. The phenotype of the mouse tumours is very immature and different from that of the T-ALL that occurred in the three X-SCID patients. Unfortunately, Woods *et al.* do not indicate whether the tumours were clonal, whether they expressed *IL2RG*, or whether JAK3 was activated; it is possible that the *IL2RG* gene might itself be mutated — for example, as a result of errors in lentiviral reverse transcription, but details of the insertion sites recovered are not given.

In our view, the retroviral-mediated expression of *IL2RG* in haematopoietic precursors does not represent a leukaemogenic event. This γ -chain is highly expressed in normal haematopoietic CD34⁺ precursors and during all stages of T-cell development, whereas *LMO2* is only expressed in progenitor cells and the earliest thymocytes⁶. Furthermore, *IL2RG* expression in treated X-SCID patients is within the

normal range⁵. Persistent activation of JAK3 was not detected in these patients, indicating that *IL2RG* function was normal in the setting of retroviral expression⁵.

Taken together, these observations argue against *IL2RG* acting as an oncogene in human gene therapy, although our experiments cannot rule out the possibility that a therapeutic transgene might provide an inappropriate growth stimulus when expressed at high dose or at an inappropriate time. Instead, it is likely that restored *IL2RG* expression allows T-cell development to progress to stages at which *LMO2* would normally be completely downregulated but might contribute to leukaemogenesis if ectopically expressed.

Karin Pike-Overzet*, Dick de Ridder†, Floor Weerkamp*, Miranda R. M. Baert*, Monique M. Versteegen‡, Martijn H. Brugman‡, Steven J. Howe§, Marcel J. T. Reinders†, Adrian J. Thrasher§, Gerard Wagemaker‡, Jacques J. M. van Dongen*, Frank J. T. Staal*

*Department of Immunology and ‡Department of Hematology, Erasmus University Medical Center, 3015 GE Rotterdam, The Netherlands
e-mail: f.staal@erasmusmc.nl

†Information and Communication Theory Group, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2628 CD Delft, The Netherlands

§Molecular Immunology Unit, Institute of Child Health, University College London, London WC1N 1EH, UK

1. Woods, N.-B., Bottero, V., Schmidt, M., von Kalle, C. & Verma, I. M. *Nature* **440**, 1123 (2006).
2. Hacein-Bey-Abina, S. *et al.* *N. Engl. J. Med.* **348**, 255–256 (2003).
3. Marshall, E. *Science* **299**, 320 (2003).
4. Check, E. *Nature* **433**, 561 (2005).
5. Hacein-Bey-Abina, S. *et al.* *Science* **302**, 415–419 (2003).
6. Dik, W. A. *et al.* *J. Exp. Med.* **201**, 1715–1723 (2005).
7. Weerkamp, F., Pike-Overzet, K. & Staal, F. J. T. *Trends Immunol.* **27**, 125–131 (2006).

doi:10.1038/nature05218

GENE THERAPY

X-SCID transgene leukaemogenicity

Arising from: N.-B. Woods, V. Bottero, M. Schmidt, C. von Kalle & I. M. Verma *Nature* **440**, 1123 (2006)

Gene therapy has been remarkably effective for the immunological reconstitution of patients with severe combined immune deficiency^{1–3}, but the occurrence of leukaemia in a few patients has stimulated debate about the safety of the procedure and the mechanisms of leukaemogenesis⁴. Woods *et al.*⁵ forced

high expression of the corrective therapeutic gene *IL2RG*, which encodes the γ -chain of the interleukin-2 receptor, in a mouse model of the disease and found that tumours appeared in a proportion of cases. Here we show that transgenic *IL2RG* does not necessarily have potent intrinsic oncogenic properties, and argue that

the interpretation of this observation with respect to human trials is overstated.

The occurrence of lymphoproliferative disease in 3 out of a total of 20 patients with X-linked severe combined immune deficiency (X-SCID), treated in two separate gene-therapy trials, was surprising as it was not predicted from prior preclinical studies in mice. However, insertional mutagenesis leading to activation of proto-oncogenic transcription factors, such as LMO2, has since been shown to be important tumour-promoting events⁶.

The possibility that *IL2RG* might itself be a contributor to oncogenesis was proposed after coincident retroviral insertion sites were found at the *IL2RG* and *LMO2* loci in a mouse tumour⁷. Woods *et al.*⁵ show that 33% (5 out of 15) of C57BL6 X-SCID mice develop T-cell lymphomas after reconstitution with either X-SCID or wild-type bone-marrow stem cells transduced with lentiviral vectors encoding murine *IL2RG*; however, X-SCID mice engrafted with wild-type bone-marrow cells treated by mock transduction, or with a potentially neutral vector encoding a reporter gene, did not develop lymphomas (up to 1.5 years' follow-up). As there were no common genomic targets in the five mice with lymphoma and because the controls did not develop disease, the authors conclude that the lymphomas were not caused by insertional mutagenesis and that the therapeutic transgene itself is intrinsically oncogenic.

We challenge this interpretation on several grounds. First, molecular data are not presented on insertion sites, although the transgene copy number is high (which, in the context of a potent, ubiquitously active promoter, will increase the risk of inadvertent mutagenesis⁸): there is insufficient information to judge whether there are significant differences in copy number and numbers of infused cells between the experimental groups, and therefore insertional mutagenesis cannot be excluded. Second, C57BL6 mice are intrinsically prone to haematological malignancy. Third, analysis of the tumours themselves was incomplete: it is unclear whether they were clonal or oligoclonal, why one mouse with a thymic lymphoma did not contain vector, and whether transgene insertion in tumour cells resulted in upregulation of recognized oncogenes (which may be located at very large distances from the insertion site).

In addition, the implications of the findings of Woods *et al.* are overstated with respect to human gene-therapy trials for X-SCID. Their transgene is overexpressed to artificially high levels (several-fold) in murine bone-marrow and spleen samples, whereas this is not the case in human studies (or in leukaemic clones derived from patients). There is no evidence in human studies for constitutive activation of JAK3-dependent signalling, indicating that *IL2RG* is functioning normally in this context (Woods *et al.*⁵ do not investigate this crucial point.)

We have tested the efficacy and toxicity of *IL2RG*-encoding transgenes in murine systems. Of 68 mice (26 X-SCID and 42 wild-type mice) reconstituted by gammaretroviral or lentiviral gene transfer and monitored for more than 6 months, only 3 cases of lymphoma were identified (refs 9, 10, and previously unpublished results). Transgene sequences were found in the tumour mass in one of these, although the relation of tumour development to gene transfer was not investigated. Of the other two tumours, one was transgene-negative and the other was not evaluated (the tissue was inadvertently discarded before it could be analysed). Finally, of two newly generated transgenic lines in which human *IL2RG* is expressed from a *CD2* promoter in wild-type mice, 54 out of 54 mice have survived for more than 12 months without developing tumours.

Our results therefore undermine the conclusions of Woods *et al.*⁵ and do not support the contention that *IL2RG* itself is potentially oncogenic. Clinical gene-therapy trials are scrutinized by the regulatory authorities and by the public, so any data pertaining to these need to be rigorously underpinned before they can be properly informative.

Adrian J. Thrasher^{1,2}, H. Bobby Gaspar^{1,2}, Christopher Baum^{3,4}, Ute Modlich³, Axel Schambach³, Fabio Candotti⁵, Makoto Otsu⁶, Brian Sorrentino⁷, Linda Scobie⁸, Ewan Cameron⁸, Karen Blyth⁸, Jim Neil⁸, Salima Hacein-Bey Abina^{9,10}, Marina Cavazzana-Calvo^{9,10,11}, Alain Fischer^{9,11,12}

¹Molecular Immunology Unit, Institute of Child Health, University College London, London WC1N 1EH, UK

e-mail: a.thrasher@ich.ucl.ac.uk

²Great Ormond Street Hospital for Children NHS Trust, London WC1N 3JH, UK

³Department of Experimental Hematology, Hannover Medical School, 30625 Hannover, Germany

⁴Division of Experimental Hematology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229-3039, USA

⁵Genetics and Molecular Biology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

⁶Center for Experimental Medicine, University of Tokyo, Tokyo 108-8639, Japan

⁷Department of Hematology/Oncology, Division of Experimental Hematology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA

⁸Division of Pathological Sciences, Institute of Comparative Medicine, Faculty of Veterinary Medicine, University of Glasgow, Glasgow G61 1QH, UK

⁹INSERM Unit 768, Hôpital Necker, ¹⁰Département de Biothérapies, ¹²Unité d'Immuno-Hématologie Pédiatrique, Assistance Publique-Hôpitaux de Paris, 75743 Paris, France

¹¹Faculté de Médecine, Université René Descartes, 75270 Paris, France

METHODS

Bone-marrow cells from X-SCID or wild-type mice (C57BL6) were transduced with replication-defective retroviral vectors encoding *IL2RG* complementary DNA and engrafted into irradiated (γ -chain-deficient; $n = 26$) or wild-type ($n = 42$) recipients. Recipient animals comprised the following overlapping experimental groups: murine *IL2RG* transgene ($n = 22$); human *IL2RG* transgene ($n = 46$); gammaretroviral vector with intact long terminal repeats ($n = 56$); lentiviral vector (internal spleen focus forming virus promoter; $n = 12$); wild-type bone marrow ($n = 42$); X-SCID bone marrow ($n = 26$). All mice were analysed periodically between 6 and 12 months after engraftment for immunophenotypic and histological abnormalities. Additionally, bone marrow from three successfully treated X-SCID animals (gammaretroviral vector, human *IL2RG*) was transplanted into a group of eight secondary recipients and followed for up to 6 months, during which time no tumours developed (not included in totals). For generation of transgenic lines, a human *IL2RG* construct (CD2 promoter) was microinjected into C57BL6 $\times$ CBA/Ca F2 fertilized eggs. Two founder animals with a confirmed integrated transgene were used to generate two independent lines. Further details of vectors and methodology are available from the authors.

1. Cavazzana-Calvo, M. *et al.* *Science* **288**, 669–672 (2000).
2. Aiuti, A. *et al.* *Science* **296**, 2410–2413 (2002).
3. Gaspar, H. B. *et al.* *Lancet* **364**, 2181–2187 (2004).
4. Hacein-Bey-Abina, S. *et al.* *N. Engl. J. Med.* **348**, 255–256 (2003).
5. Woods, N.-B., Bottero, V., Schmidt, M., von Kalle, C. & Verma, I. M. *Nature* **440**, 1123 (2006).
6. Hacein-Bey-Abina, S. *et al.* *Science* **302**, 415–419 (2003).
7. Dave, U. P., Jenkins, N. A. & Copeland, N. G. *Science* **303**, 333 (2004).
8. Modlich, U. *et al.* *Blood* **105**, 4235–4246 (2005).
9. Soudais, C. *et al.* *Blood* **95**, 3071–3077 (2000).
10. Otsu, M. *et al.* *Mol. Ther.* **1**, 145–153 (2000).

doi:10.1038/nature05219

GENE THERAPY

Woods *et al.* reply

Replying to: K. Pike-Overzet *et al.* *Nature* **443**, doi:10.1038/nature05218 (2006); A. J. Thrasher *et al.* *Nature* **443**, doi:10.1038/nature05219 (2006)

Using a lentiviral vector and a murine model of X-linked severe combined immune deficiency (X-SCID), we have shown that mice transplanted with cells constitutively overexpressing the therapeutic gene *IL2RG*, a subunit found in

several cytokine receptors, can develop thymic lymphomas 6 to 12 months later¹. Our results contradict those of Pike-Overzet *et al.*² and Thrasher *et al.*³.

We do not claim that insertional mutagenesis

plays no part in this oncogenesis. We have analysed integration sites of lymphomic *IL2RG*-transduced mice tissues¹ and found dominant clones in thymuses that have integration profiles similar to those seen in the spleen and/or bone marrow of the individual animals; this suggests that the lymphoma is clonal. However, we found no common integration sites among the mice, as was seen in the X-SCID patients affected in the gene-therapy trial, where three out of three patients had insertional gene activation of the *LMO2* gene locus. Also, as none of the mice originally transplanted with a lentiviral control vector developed lymphomas¹, we maintain that insertional mutagenesis was not the only oncogenic event. Moreover, a lentiviral vector construct similar to the one we used has recently been shown to be significantly less prone to causing insertional mutagenesis than the oncoretroviral vectors that drive expression from the long terminal repeats⁴.

IL2RG transgene expression was high in our mice that developed lymphomas¹, with even the lowest amount of expression being several-fold higher than in normal transplanted bone marrow. Previously, lymphomas in X-SCID-treated mice were rarely found, probably because of the short duration of studies. And oncoretroviral expression of *IL2RG* was on average normal to subnormal^{5–9}, unlike in our transplanted mice (33% of which developed lymphomas)¹. In secondary mice transplanted from three primary retrovirus transgene-treated donors, expression of *IL2RG* was below endogenous levels⁵. Although the discrepancy with our results¹ may be ascribed to our much longer follow-up period (up to 1.5 years), it is possible that lymphoma development may have been

linked with the increased expression of *IL2RG* in our mice.

Our results should be interpreted with caution and not be extrapolated into drastically changing or terminating currently successful gene-therapy protocols. If increased expression of *IL2RG* is potentially oncogenic in mice, an additional safety measure would be to ensure that *IL2RG* transgene expression is physiological. The mouse model may not necessarily reflect the patient setting, however, as there are significant differences between mice and humans in signalling by some of the *IL2RG*-containing interleukin receptors. *IL2RG* signals through the activation of the JAK3 molecule and we are currently analysing samples of lymphoma tissue from our mice for the presence of activated JAK3, which would provide further evidence that *IL2RG* in this context is contributing to the development of the tumours. In two of the three patients who developed lymphoma after gene therapy, however, there was no constitutive activation of JAK3 in their tumours.

Clinical experience in treating X-SCID patients with retroviral vectors that express *IL2RG* is increasing. Although a significant number of *IL2RG*-expressing T-cell clones exist in each treated patient because each of their T cells is transgenic, more than 5 years of follow-up have been uneventful in several cases. Also, no *IL2RG*-related lymphoproliferation has been reported in humans in the absence of *LMO2* activation. However, oncoretroviral vector expression from the long terminal repeat is subject to positional variegation, and effects on transgene expression can occur that are dependent on the site of integration. For

example, a therapeutic retrovirus vector might land in a chromosomal site where expression of the transgene is enhanced, leading to over-expression.

Our intention was to warn other investigators that the *IL2RG* gene, when overexpressed in some settings, has the potential to be oncogenic — a finding that is not altogether surprising for a transgene encompassing the signalling portion of a growth-factor receptor. Further details, including the clonality of these lymphomas and quantification of the expression of *IL2RG*, and analysis of downstream signalling pathways, will shortly be submitted for publication.

Niels-Bjarne Woods*, **Virginie Bottero***,
Manfred Schmidt†, **Christof von Kalle†**,
Inder M. Verma*

*Laboratory of Genetics, The Salk Institute for Biological Studies, La Jolla, California 92037, USA
e-mail: verma@salk.edu

†National Center for Tumor Diseases Heidelberg and Department of Translational Oncology, German Cancer Research Center, 69120 Heidelberg, Germany

1. Woods, N.-B., Bottero, V., Schmidt, M., von Kalle, C. & Verma, I. M. *Nature* **440**, 1123 (2006).
2. Pike-Overzet, K. et al. *Nature* **443**, doi:10.1038/nature05218 (2006).
3. Thrasher, A. J. et al. *Nature* **443**, doi:10.1038/nature05219 (2006).
4. Montini, E. et al. *Nature Biotechnol.* **24**, 687–696 (2006).
5. Soudais, C. et al. *Blood* **95**, 3071–3077 (2000).
6. Lo, M. et al. *Blood* **94**, 3027–3036 (1999).
7. Otsu, M. et al. *Mol. Ther.* **1**, 145–153 (2000).
8. Otsu, M., Sugamura, K. & Candotti, F. *Blood* **97**, 1618–1624 (2001).
9. Aviles Mendoza, G. J. et al. *Mol. Ther.* **3**, 565–573 (2001).

doi:10.1038/nature05220

Central nervous system control of food intake and body weight

G. J. Morton¹, D. E. Cummings², D. G. Baskin^{2,3}, G. S. Barsh⁴ & M. W. Schwartz¹

The capacity to adjust food intake in response to changing energy requirements is essential for survival. Recent progress has provided an insight into the molecular, cellular and behavioural mechanisms that link changes of body fat stores to adaptive adjustments of feeding behaviour. The physiological importance of this homeostatic control system is highlighted by the severe obesity that results from dysfunction of any of several of its key components. This new information provides a biological context within which to consider the global obesity epidemic and identifies numerous potential avenues for therapeutic intervention and future research.

The apparent ease with which we decide whether or not to eat an appetizing food testifies to the efficiency with which the central nervous system (CNS) processes information of surprising variety and complexity. With the aid of cognitive, visual and olfactory cues, food items must first be identified and distinguished from a nearly infinite array of potentially toxic environmental constituents. Using taste information, the food's palatability is then assessed and integrated with both short- and long-term signals regarding nutritional state. One consequence of this integration is that the drive to eat decreases as food is ingested (termed 'satiation'), ensuring that the amount consumed in a single meal does not exceed what the body can safely handle. Changing energy requirements is another factor that can influence food consumption. Through a process known as energy homeostasis, food intake is adjusted over time so as to promote stability in the amount of body fuel stored as fat. In this way, through diverse blood-borne and afferent neural signals, information regarding nutrient status and energy stores is communicated to the brain where it is integrated with cognitive, visual, olfactory and taste cues—all happening unconsciously, before the first bite is taken.

Here we describe CNS mechanisms that regulate food intake, and review evidence that in response to reduced body fat stores, adaptive changes occur in neuronal systems governing both food-seeking behaviour (important for meal initiation) and satiety perception (important for meal termination). The net effect is that in response to weight loss, both the motivation to find food and the size of individual meals tend to increase until energy stores are replenished (Fig. 1), and mutation of any of several key molecules involved in this process has been shown to cause severe obesity in both animal models and humans. Despite this progress, the many fundamental questions remaining unanswered represent rich opportunities for future study.

Energy homeostasis

Obesity, by definition, results from ingesting calories in excess of ongoing requirements. Although environmental and lifestyle factors contribute to obesity pathogenesis, homeostatic adaptations to weight loss induced by voluntary caloric restriction are robust in both lean and obese individuals. In addition, normal-weight individuals are protected against expansion of body fat stores induced by

overfeeding^{1,2}, indicating that biological mechanisms protect against weight gain as well as weight loss, at least in normal-weight individuals. Together, these findings indicate that obesity involves the defence of an elevated body weight, rather than the absence of regulation, and that deleterious interactions between obesity-promoting environmental factors and homeostatic control systems contribute to common forms of obesity and, hence, the global obesity pandemic.

Adiposity negative feedback. Introduced more than 50 years ago, the 'adiposity negative-feedback' model of energy homeostasis is founded on the premise that circulating signals inform the brain of changes in body fat mass and that in response to this input, the brain mounts adaptive adjustments of energy balance to stabilize fat stores³. Proposed criteria for a negative-feedback signal include: (1) that it circulates at levels proportionate to body fat content and enters the brain; (2) that it promotes weight loss by acting on neuronal systems implicated in energy homeostasis; and (3) that blockade of these neuronal actions increases food intake and body weight. Although many nutrients (for example, free fatty acids and glucose), cytokines (for example, interleukin-6, tumour necrosis factor- α) and hormones (for example, glucocorticoids) fulfill some of these criteria, only leptin and insulin satisfy all of them⁴.

Studies in primitive organisms such as the nematode, *Caenorhabditis elegans*, and the fruitfly, *Drosophila melanogaster*, implicate insulin as a key ancestral negative-feedback regulator of body fuel stores^{5,6}. By comparison, leptin has not been detected in invertebrates and probably evolved more recently⁷. Although genetic and pharmacological studies^{8,9} suggest a more critical role for leptin than insulin in mammalian energy homeostasis, cross-talk between these hormones with respect to both the neuronal subsets and signal transduction pathways on which they act offers evidence of their shared evolutionary past.

Although leptin administration causes weight loss in diverse mammalian species, enthusiasm surrounding leptin as a therapeutic agent diminished rapidly with the discovery that leptin resistance is common among obese individuals¹⁰. Because obesity has long been associated with insulin resistance in peripheral tissues, it is perhaps not surprising that in obese rats, the hypothalamus develops resistance to insulin¹¹ as well as leptin¹². Although reduced neuronal signalling by either hormone induces hyperphagia and weight gain

¹Department of Medicine, Harborview Medical Center and University of Washington, Seattle, Washington 98104, USA. ²Department of Medicine, Veterans Affairs Puget Sound Health Care System and University of Washington, Seattle, Washington 98108, USA. ³Departments of Medicine and Biological Structure, University of Washington, Seattle, Washington 98104, USA. ⁴Departments of Genetics and Pediatrics, Stanford University School of Medicine, Stanford, California 94305, USA.

in animal models^{13,14}, whether neuronal resistance to these hormones contributes to the pathogenesis of common obesity is an important unanswered question.

Gastrointestinal hormones. Peptide YY_{3–36} (PYY_{3–36}) is an enteric hormone that can reduce food intake in rodents and humans¹⁵. Whereas plasma levels of PYY_{3–36} decline in advance of meals, levels of the gastric hormone ghrelin rise shortly before meals and fall sharply on feeding¹⁶. Opposing the hypothalamic actions of leptin, insulin and PYY_{3–36}, ghrelin powerfully stimulates food intake in multiple species, including humans^{17–19}, suggesting that the effect of weight loss to increase ghrelin levels²⁰ may (along with reduced leptin and insulin levels) contribute to weight regain.

Nutrient-related signals. In addition to molecular targets of current anti-obesity drugs, including receptors for endocannabinoids, norepinephrine and serotonin^{21–23}, several nutrient-related signals are implicated in the homeostatic control of feeding. Among these are free fatty acids, which exert insulin-like effects in key brain areas for energy homeostasis, including the hypothalamic arcuate nucleus (ARC), possibly by favouring intracellular accumulation of long-chain fatty acyl-CoA (LCFA-CoA)^{24,25}. The hypothesis that hypothalamic LCFA-CoA signalling has a critical role in energy homeostasis is supported by the obesity that results when the level of these molecules is selectively reduced in rat ARC²⁶.

Whereas intracellular LCFA-CoA accumulation is proposed to signal nutrient abundance, the enzyme AMP-activated protein kinase (AMPK) is a sensor of nutrient insufficiency. When cells experience a critical drop of fuel availability (as reflected by an increased AMP/ATP ratio), AMPK activation increases substrate oxidation to replenish depleted ATP levels. In mediobasal hypothalamus, activation of AMPK increases food intake and body weight whereas conversely, inhibition of AMPK has the opposite effect²⁷. Because hypothalamic AMPK activity is inhibited by leptin and insulin²⁷, but stimulated by ghrelin²⁸, altered AMPK signalling may contribute to the feeding effects of these hormones. One consequence of increased AMPK activity is oxidation of LCFA-CoA, and this mechanism might contribute to its orexigenic effects. AMPK also decreases the activity of another nutrient-sensing enzyme, mammalian target of rapamycin (mTOR), implicated in hypothalamic leptin action²⁹.

Regulation of satiety perception

Most vertebrates consume food in discrete bouts or ‘meals’. Although numerous environmental variables can influence meal initiation, satiety perception is a more biologically determined and predictable process³⁰, triggered on nutrient ingestion by gastric distension and the release of gut factors including cholecystokinin (CCK)³¹. Many such ‘satiety signals’ are transmitted to the brain via vagal afferent fibres that synapse in the nucleus tractus solitarius (NTS) in the hindbrain, which participates in gustatory, satiety and visceral sensation.

Because homeostatic adjustments of food consumption must, by definition, involve changes of meal size, meal frequency, or both, systems that control energy homeostasis must integrate with those governing intake on a meal-to-meal basis. In rats, leptin reduces intake, at least in part, by enhancing the response to satiety signals³², whereas conversely, obese, leptin-receptor-deficient rats exhibit blunted satiety responses to CCK, and this defect is ameliorated by restoring leptin receptors to the ARC³³. This result suggests that in response to afferent input from leptin, homeostatic circuits in the hypothalamus elicit compensatory reductions of food intake by enhancing the response to meal-related satiety signals. Brain areas beyond the ARC also probably contribute to leptin’s enhancement of the response to satiety signals, because leptin receptors are present in many brain areas involved in food intake control, including the NTS itself^{34,35}, and because leptin administration directly into the NTS reduces food intake³⁴. The integration of long-acting homeostatic and short-acting satiety signals may therefore involve direct actions of leptin on NTS neurons that process input from vagal afferent fibres, in addition to its effects on neurons in hypothalamus and elsewhere that project to the NTS (Fig. 2)³⁶.

Mechanisms of food reward

Perception of food reward begins with information generated by oral taste receptor cells that is subsequently transmitted to the NTS by afferent sensory fibres. From the NTS, taste information is conveyed to multiple sites in the hindbrain (for example, parabrachial nucleus), midbrain (ventral tegmental area or VTA) and forebrain (for example, nucleus accumbens (NAc), striatum, thalamus and cerebral cortex)³⁷, which collectively sense and discriminate among

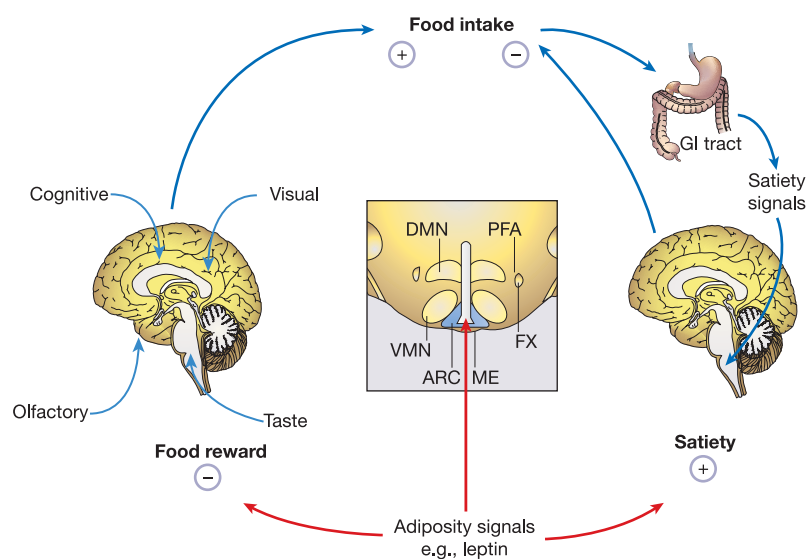


Figure 1 | Model for negative-feedback regulation of food intake in response to changes in body fat content. Hormones such as leptin circulate in proportion to body fat mass, enter the brain, and act on neurocircuits that govern food intake. Through both direct and indirect actions, leptin diminishes perception of food reward—the palatability of food—while enhancing the response to satiety signals generated during food

consumption that inhibit feeding and lead to meal termination. The effect of weight loss to lower leptin levels (and hence to reduce leptin signalling) increases rewarding properties of food while diminishing satiety, a combination that potentially increases intake. ARC, arcuate nucleus (highlighted in blue); DMN, dorsomedial nucleus; FX, fornix; ME, median eminence; PFA, perifornical area; VMN, ventromedial nucleus.

different tastes and textures, assigning reward value to them (Fig. 3). Higher processing of taste information in primates involves 'primary taste neurons' in the insular cortex that are supplied by NTS neurons by way of the thalamus. These cells project in turn to 'secondary taste neurons' in the orbitofrontal cortex that integrate taste information with relevant olfactory, visual and cognitive inputs³⁸. The response of these latter cells (but not of primary taste neurons) to taste stimuli is described as 'hunger-dependent', in that it decreases as food is eaten, implying the capacity to integrate taste information with satiety-related (and perhaps adiposity-related) input³⁸; data obtained using functional magnetic resonance imaging in humans support this view³⁹. According to this hypothesis, reduced firing of secondary taste neurons diminishes the reward value of foods and thereby contributes to meal termination via projections to the striatum and amygdala, areas that attach motivational value needed to engage motor outputs³⁹. Whether activity of secondary taste neurons in the orbitofrontal cortex is sensitive to input from adiposity and satiety signals, and whether it is either necessary or sufficient to explain hedonic stimulation of feeding behaviour, are questions that warrant further study.

Reward valuation involves the release of dopamine from neurons that originate in the VTA and project to NAc, striatum and other brain areas^{37,40}. Acting in these forebrain areas, dopamine potently augments the drive to obtain a rewarding stimulus (that is, increases the 'wanting' of a particular food or drug)⁴⁰, but it is not directly responsible for the hedonic experience itself (that is, the 'liking' of a palatable food). Rather, μ -opioid receptor signalling in the NAc and adjacent forebrain structures (activated in part by dopamine release) is implicated in the hedonic experience^{37,40}, although wanting and liking ordinarily occur together.

One mechanism whereby activation of the VTA $\rightarrow$ NAC pathway may promote consumption of palatable food involves projections to the lateral hypothalamic area. This brain area contains neurons that potentially stimulate food intake and is supplied by fibres not only from striatum and orbitofrontal cortex, but also from the ARC

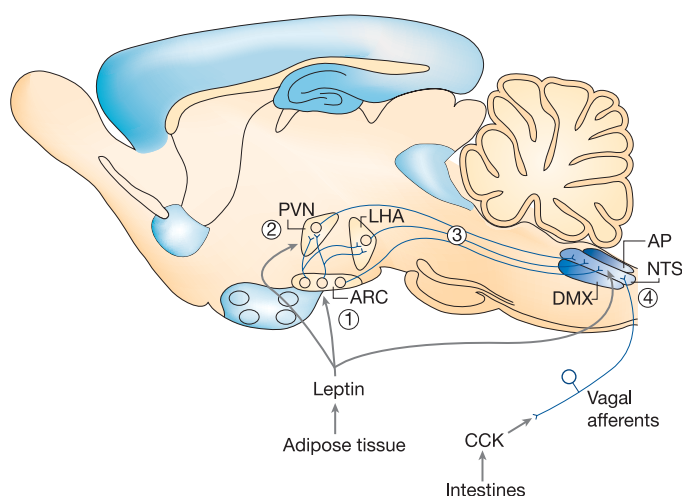


Figure 2 | Model for integration of adiposity- and satiety-related inputs. Afferent input in proportion to body fat mass (for example, leptin) enhances the response to satiety signals such as CCK that are generated in response to food consumption and lead to meal termination. Whereas the hypothalamus is a key target for leptin action, the satiety effect of CCK involves activation of vagal afferent fibres that terminate in the NTS. Integration of these inputs can involve the actions of leptin in the ARC (1) or other hypothalamic areas (2) involving neurons that project to the NTS (3) and influence the response to CCK (4). In addition, leptin can act directly in the NTS. AP, area postrema; DMX, dorsal motor nucleus of the vagus nerve; LHA, lateral hypothalamic area; NTS, nucleus of the solitary tract; PVN, paraventricular nucleus.

and other crucial hypothalamic areas for energy homeostasis. Food-intake-stimulatory neurons in the lateral hypothalamic area seem to be constrained by tonic inhibition that can be relieved by activation of reward pathways, thereby engaging motor programs that stimulate feeding behaviour³⁷. Lateral hypothalamic area neurons supplying the NTS may, in addition, attenuate the response to satiety signals, increasing the amount of food consumed during a meal. These considerations support a view of the lateral hypothalamic area as an integrative node for homeostatic, satiety and reward-related inputs that collectively govern motor programs that activate feeding behaviour (Fig. 3).

Regulation of brain reward circuitry

The concept that reward perception is subject to homeostatic regulation derives from evidence that food deprivation strongly augments the reward value (for example, the speed of learning to obtain a rewarding stimulus, the dose of the stimulus needed to establish its rewarding value, or the amount of work an animal will perform to obtain the stimulus) of addictive drugs including heroin, amphetamine and cocaine^{41,42}. As fasting also augments the reinforcing value of electrical stimulation of brain ‘pleasure centres’⁴³, reduced food availability seems to exert a global, stimulatory effect on reward perception. One mechanism to explain this effect proposes that leptin and insulin tonically inhibit brain reward circuitry and that, by lowering circulating levels of these hormones,

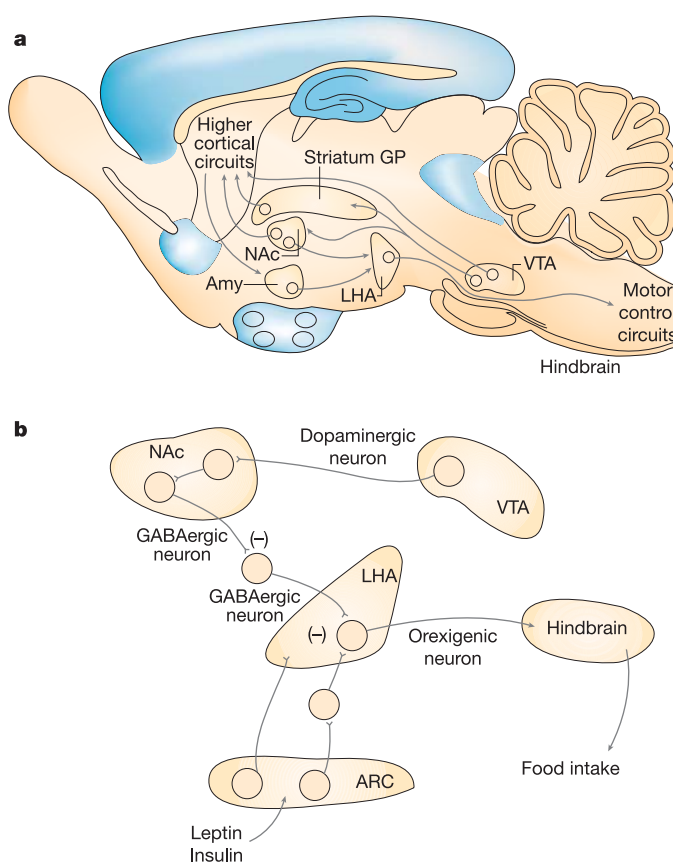


Figure 3 | Model for integration of adiposity- and reward-related inputs. Brain reward circuitry. **a**, Neurons in the VTA of the midbrain project to forebrain areas including the NAc, striatum and cortex, and assign reward value to palatable food. **b**, Perception of pleasure associated with consumption of a palatable food involves neuronal activation in the NAc and striatum, which through activation of opiate peptide receptors disinhibits the lateral hypothalamic area and thereby stimulates feeding. Amy, amygdala; GP, globus pallidus; NAc, nucleus accumbens; VTA, ventral tegmental area.

energy restriction increases the sensitivity of reward circuits^{43,44}. Consistent with this hypothesis, centrally administered insulin or leptin diminishes both sucrose preference (a measure of food reward)⁴⁴ and the effect of fasting to increase the rewarding properties of electrical pleasure-centre stimulation⁴³. At the cellular level, increased expression of dopamine re-uptake transporters in VTA neurons, which reduces synaptic dopamine levels in the NAc⁴⁵, may contribute to the inhibitory effect of adiposity-related hormones on food reward. Collectively, these observations suggest that by decreasing neuronal input from adiposity signals, energy restriction increases responses to rewarding stimuli as an adaptive mechanism motivating animals threatened by caloric insufficiency to seek and obtain palatable foods.

Neuronal sensing of adiposity-related signals

Recent studies have yielded testable, molecular models⁴⁶ to investigate how adiposity signals regulate the function of hypothalamic neurons. Like the ARC, the hypothalamic ventromedial nucleus was recently shown to be essential for leptin regulation of energy balance⁴⁷, and identifying leptin-sensitive neurons in this area has become a high priority. Meanwhile, substantial progress has been made towards understanding how ARC neurons participate in this process. Among insulin- and leptin-sensing ARC neurons are those that co-express neuropeptide Y (NPY) and agouti-related protein (AgRP), neuropeptides that stimulate food intake via distinct mechanisms. These NPY/AgRP neurons are inhibited by leptin, insulin and PYY₃₋₃₆, whereas they are stimulated by ghrelin⁴⁶. Consequently, weight loss activates these neurons through a combination of reduced inhibitory and increased stimulatory input⁴⁶ (Fig. 4a). Adjacent to these cells are neurons that express pro-opiomelanocortin (POMC), the polypeptide precursor from which melanocortins such as α -melanocyte stimulating hormone (α -MSH) are derived. On release from axon terminals, α -MSH binds to and activates neuronal melanocortin receptors, thereby decreasing food intake and favouring weight loss⁴⁸. POMC neurons are stimulated by leptin⁴⁹ but inhibited by neighbouring NPY/AgRP neurons⁴⁹. In response to

weight loss, activation of NPY/AgRP neurons is coupled to inhibition of POMC neurons, a combination that favours the recovery of lost weight and reflects the integration of diverse humoral and neuronal inputs⁴⁶.

Numerous observations support this model of how energy homeostasis circuitry is organized and regulated. For example, absence of leptin in obese (*Lep^{ob}/Lep^{ob}*) mice reduces *Pomc* expression in the ARC⁵⁰ while increasing levels of NPY and AgRP^{4,51}, mimicking the hypothalamic response to starvation. Similarly, absence of α -MSH in *Pomc* knockout mice causes hyperphagia and obesity⁵², and blockade of neuronal melanocortin signalling diminishes the response to central leptin administration⁵³. By comparison, deletion of *Npy*, *AgRP*, or both genes causes only mild defects in energy balance^{54,55}, raising questions about the relative importance of orexigenic versus anorexigenic signals in food intake and body weight regulation.

An oft-cited explanation for when knockout genotypes fail to yield abnormal phenotypes is 'genetic redundancy', in which loss of one gene is somehow compensated by altered expression of other genes or through more complex homeostatic mechanisms. How, then, are we to interpret the mild phenotype observed in mice lacking NPY and AgRP? To answer this question, investigators have recently ablated NPY/AgRP neurons themselves (rather than the genes encoding *Npy* and *AgRP*). Notably, the outcome of these studies depends on both the age of the animal and the time course over which the ablation occurs. In two studies, rapid ablation was accomplished by expressing the human receptor for diphtheria toxin selectively in NPY/AgRP neurons and injecting those animals with diphtheria toxin. (Humans have a gene that encodes a receptor for diphtheria toxin but mice do not; thus administration of low doses of diphtheria toxin to mice is normally harmless.) Using this approach, rapid ablation of NPY/AgRP neurons induced profound, life-threatening anorexia in adult mice^{56,57}, but not in mice in which these neurons were ablated during the neonatal period⁵⁶. By contrast, gradual ablation over a period of months—brought about either by expressing a polyglutamine-containing protein⁵⁸ or by removing the capacity for oxidative

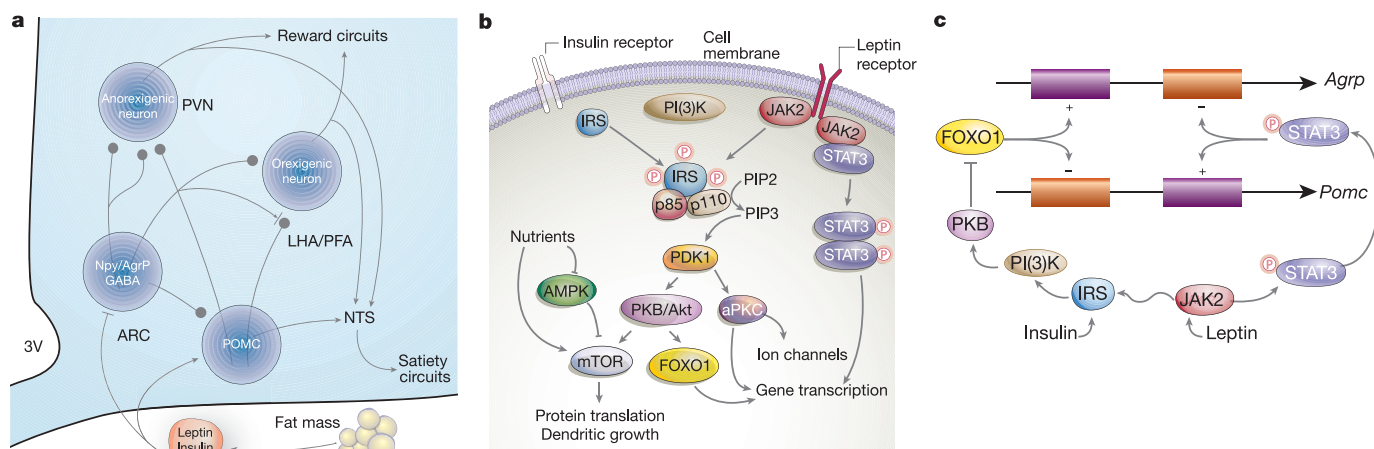


Figure 4 | Hypothalamic neurocircuits and signal transduction mechanisms involved in energy homeostasis. **a**, The arcuate nucleus (ARC) contains both NPY/AgRP neurons, which are inhibited by insulin and leptin and, when activated, stimulate food intake, and POMC neurons, which reduce food intake and are stimulated by insulin and leptin. Both neuronal subsets project to adjacent hypothalamic areas including the paraventricular nucleus (PVN), where neurons that reduce food intake are concentrated (anorexigenic), and the lateral hypothalamic area (LHA), which contains neurons that stimulate food intake (orexigenic). Outflow from each of these hypothalamic nuclei supplies 'downstream' brain areas involved in the perception of satiety (including the nucleus of the solitary tract (NTS) in the hindbrain) as well as food reward. NPY/AgRP neurons also inhibit POMC

neurons via synaptic release of the inhibitory transmitter, GABA (γ -aminobutyric acid). 3V, third cerebral ventricle; PFA, perifornical area. **b**, On binding to their respective receptors, both insulin and leptin activate the IRS–PI(3)K pathway. Phosphorylation of IRS proteins on tyrosine residues is induced by both the insulin receptor and by JAK2, a tyrosine kinase coupled to the leptin receptor, and allows IRS proteins to activate PI(3)K. By converting PIP2 to PIP3, PI(3)K activates PDK1, which in turn activates downstream enzyme systems including PKB and atypical PKC. Leptin receptor activation of JAK2 also induces tyrosine phosphorylation of STAT3, which dimerizes to become a transcription factor. **c**, Model for reciprocal regulation of *Pomc* and *AgRP* gene transcription via FOXO1 (which is inhibited by PKB) and STAT3 (modified from ref. 67).

phosphorylation specifically in NPY/AgRP neurons⁵⁹—had only mild effects on energy balance. Thus, orexigenic signals from these neurons have a critical role in body weight regulation, but if loss of these cells occurs early in life or via a gradual, progressive process, survival is ensured by as-yet-unidentified compensatory mechanisms.

What kinds of molecular and cellular mechanisms might underlie homeostatic compensation for NPY/AgRP neurodegeneration? An intriguing indication comes from studies using ciliary neurotrophic factor, which activates many of the same downstream signalling molecules as leptin, but which is not subject to neuronal resistance seen with leptin in obese models⁶⁰. Surprisingly, ciliary neurotrophic factor stimulates proliferation of leptin-responsive neurons in the ARC⁶¹, an intervention that may alter fundamentally the balance of orexigenic versus anorexigenic pathways in adult animals and thereby combat leptin resistance that accompanies obesity.

Signal transduction in ARC neurons. On insulin binding, the insulin receptor recruits several intracellular molecules involved in signal transduction⁶². Among these are insulin receptor substrate (IRS) proteins, which are phosphorylated on tyrosine residues by the activated insulin receptor. This, in turn, allows IRS proteins to activate phosphatidylinositol-3-OH kinase (PI(3)K), which generates phosphatidylinositol-3,4,5-trisphosphate (PIP3) from phosphatidylinositol-4,5-bisphosphate (PIP2). PIP3-mediated activation of PDK1 activates an enzyme cascade that includes protein kinase B (PKB, also known as Akt) and members of the atypical PKC family⁶². Among the targets of activated PKB are mTOR and the transcription factor FOXO1, which is inhibited by PKB-mediated phosphorylation (Fig. 4b).

The leptin receptor is a class 1 cytokine receptor that regulates gene transcription via activation of the Janus kinase–signal transducer and activator of transcription (JAK–STAT) pathway. Leptin binding to the ‘long’ or ‘signalling’ form of the leptin receptor (LeprB) stimulates the tyrosine kinase JAK2 to phosphorylate STAT3 at tyrosine residues⁶³. Phospho-STAT3 dimers subsequently enter the nucleus and regulate transcription of target genes. In some cells, including hypothalamic neurons⁶⁴, leptin also activates the IRS–PI(3)K pathway⁶⁵, one of several points of convergence and synergism between intracellular signalling pathways used by insulin and leptin. Together with evidence that hypothalamic IRS–PI(3)K signalling is increased by both hormones^{65,66}, convergent actions of leptin and insulin involving PI(3)K and STAT3 activation are proposed to regulate key energy homeostasis neurocircuits.

Recent analysis of transcriptional regulation of ARC neuropeptides by insulin and leptin supports this view. Consensus DNA binding sequences for both FOXO1 (a transcription factor that is inhibited by PI(3)K) and STAT3 exist close to one another in both *Agrp* and *Pomc* promoters, and the effect of STAT3 on transcription of both neuropeptide genes is opposite to that of FOXO1⁶⁷. Thus, PI(3)K activation in response to insulin is proposed to inhibit FOXO1-mediated gene transcription. This response complements the effect of leptin-mediated STAT3 activation and provides a feasible explanation for how, during weight loss or other conditions of low plasma insulin and leptin levels, hypothalamic AgRP expression increases while POMC levels decline (Fig. 4c).

To understand better how ARC neurons are regulated by insulin, leptin and other inputs, several groups have sought to measure membrane potential and firing rate, in addition to expression of neuropeptide genes. The hypothesis that PI(3)K signalling links input from insulin and leptin to changes of electrical activity in ARC neurons received early, direct support. Specifically, both hormones were reported to activate ATP-sensitive potassium (K_{ATP}) channels and thereby hyperpolarize a subset of ‘glucose-responsive’ ARC neurons via a PI(3)K-dependent mechanism⁶⁸. These K_{ATP} channels have subsequently emerged as key cellular targets for the actions of free fatty acids and other nutrient-related signals in the ARC. However, recent work suggests that although both insulin and

leptin activate PI(3)K signalling in arcuate POMC cells⁶⁴, the depolarizing effect of leptin on these cells^{49,69} is reportedly opposed by a hyperpolarizing action of insulin⁶⁹. Thus, PI(3)K signalling does not provide a unifying explanation for how these hormones affect POMC neuron firing rate and, moreover, insulin and leptin evidently have dissimilar effects on this parameter, despite their convergent effects to stimulate *Pomc* gene transcription⁶⁷.

As expected, leptin inhibits electrical activity of NPY/AgRP neurons^{69,70}, but whereas insulin increases PI(3)K signalling in these cells, leptin has the opposite effect⁶⁴. The observation that leptin increases PI(3)K activity and firing rate in POMC neurons while inhibiting both parameters in NPY/AgRP neurons supports a causal link between leptin regulation of PI(3)K and neuronal activity, but this link does not hold for insulin. Whether and how PI(3)K activation affects neuronal firing, therefore, seems to vary with both the cell type and hormonal stimulus. In this context, the recent finding that leptin activates the nutrient-sensing enzyme mTOR in ARC neurons²⁹ is of interest. Because central inhibition of mTOR blocks leptin-mediated food intake suppression²⁹, mTOR activity seems to be essential for leptin action in this brain area. Yet, whereas mTOR is not implicated in the control of neuronal firing, it is a key determinant of dendritic growth and synaptic plasticity⁷¹. Because mTOR activation is well documented in response to both nutrient-related and IRS–PI(3)K signalling (Fig. 4b), it is possible that, rather than serving to link input from adiposity signals to firing rate, PI(3)K signalling in some hypothalamic neurons influences dendritic growth and synaptic input.

To clarify the involvement of specific molecules in ARC neurons *in vivo*, mice have been created in which IRS-2 (ref. 69), STAT3 or leptin receptor⁷² is selectively deleted from POMC cells. The mild or absent obesity phenotype of these animals contrasts sharply with the severe obesity that results from *Pomc* gene deletion⁵². A logical interpretation of these findings is that, although neuronal melanocortin signalling is critical for energy homeostasis, neither leptin receptor → JAK–STAT nor IRS–PI(3)K signalling is essential for proper functioning of these cells. However, an alternative explanation is suggested by evidence that NPY/AgRP neurons exert a tonic inhibitory effect on neighbouring POMC cells⁴⁹, and that these two neuronal subsets are regulated in a reciprocal manner by many of the same afferent inputs (Fig. 4a). Nonlinearity in the relationship between the regulation of ARC neurons and their effects on food intake is illustrated through the following predictions that incorporate this arrangement. Mice with POMC-cell-specific deletion of the leptin receptor can be expected to exhibit reduced basal POMC neuron activity, creating a leptin-resistant state and increasing food intake. As weight is gained, rising plasma levels of leptin and insulin should subsequently inhibit NPY/AgRP neurons, which, in turn, will relieve tonic constraint of POMC cells. Through this mechanism the consequences of a reduced leptin signalling on POMC cells are mitigated and further weight gain limited.

These considerations highlight how the integrated, nonlinear nature of energy homeostasis neurocircuitry can complicate the interpretation of seemingly straightforward experiments. As the list of neuronal cell types and signal transduction molecules involved in energy homeostasis continues to expand, a ‘systems biology’ approach will help to more accurately model outcomes when molecules are deleted in a cell-specific manner.

Synaptic plasticity

Arcuate neurons are regulated not only by humoral signals but also by local excitatory and inhibitory synapses, and the balance of these inputs seems to be affected by nutritional state. Specifically, fasting decreases excitatory, while increasing inhibitory, synaptic contacts on POMC neurons, whereas the opposite applies for NPY/AgRP neurons⁷³. Although the mechanism mediating these effects is unknown, it seems to involve the effect of fasting to lower leptin levels, as leptin-deficient mice display a similar alteration of synaptic

input. In addition, an unidentified subset of mediobasal ventromedial nucleus neurons provides excitatory synaptic input to POMC cells, the magnitude of which is also attenuated by fasting⁷⁴. Clarifying the extent to which synaptic regulation of hypothalamic neurons can influence the defended level of body adiposity, determining whether diet-induced obesity involves changes of brain circuitry at the synaptic level, and delineating the mechanisms underlying such effects are important scientific priorities. Realizing these goals requires new strategies for single-cell imaging, analysis of synaptic function, and modelling how synaptic changes affect energy homeostasis neurocircuitry and, consequently, the defended level of body weight.

Concluding remarks

Theoretically, drugs that target neuronal receptors for leptin, insulin, ghrelin, melanocortins, NPY or other relevant ligands have important potential, but hoped-for therapeutic breakthroughs have yet to emerge. Although delays between basic discovery and therapeutic application are inevitable, two obstacles that lie along this path are noteworthy. One of these is the integrated nature of energy homeostasis neuronal systems, which predicts that the efficacy of interventions targeting one neuronal subset or signal transduction pathway is inherently limited by compensatory responses elsewhere. As for many other chronic illnesses, effective prevention or treatment of obesity may therefore require drug combinations that target discrete components of energy homeostasis, satiety or food reward systems. However, drug combinations cannot be approved by the US Food and Drug Administration unless the drugs are individually proven safe and effective. Consequently, pharmaceutical companies have little incentive to pursue strategies involving drug combinations unless each of the drugs under consideration is, by itself, 'approvable'. This creates a situation in which potentially promising drug combinations are not developed because of limited efficacy of individual drugs when used alone.

Another obstacle arises from a lack of scientific consensus regarding fundamental aspects of obesity pathogenesis. During caloric restriction, there is little question that reduced neuronal input from adiposity-related hormones activates responses (increased food intake, decreased metabolic rate) that favour recovery of lost weight. Whether the reverse is also true—that increased neuronal input from these hormones protects against weight gain—is hotly debated. Until this issue is resolved, the importance of several key observations, including CNS resistance to leptin¹² and insulin¹¹ documented in common forms of obesity, will remain uncertain. This is because if adiposity negative-feedback signals do not normally protect against weight gain, neuronal resistance to insulin and leptin cannot cause obesity. According to this view, the growing obesity epidemic can be attributed to an inherent lack of protection against obesity-promoting environmental factors⁷⁵, rather than to an underlying homeostatic disorder. Alternatively, if adiposity negative-feedback signals do, in fact, confer protection against obesity in normal-weight individuals, neuronal resistance to these signals must, by definition, favour weight gain, and unravelling the underlying causes takes on both pathophysiological and therapeutic urgency. As our understanding of normal and abnormal regulation of food intake and body adiposity grows in its sophistication, overcoming these obstacles will create new opportunities for therapeutic intervention.

1. Sims, E. A. *et al.* Endocrine and metabolic effects of experimental obesity in man. *Recent Prog. Horm. Res.* **29**, 457–496 (1973).
2. Leibel, R. L., Rosenbaum, M. & Hirsch, J. Changes in energy expenditure resulting from altered body weight. *N. Engl. J. Med.* **332**, 621–628 (1995).
3. Kennedy, G. The role of depot fat in the hypothalamic control of food intake in the rat. *Proc. R. Soc. Lond. B* **140**, 578–592 (1953).
4. Schwartz, M. W., Woods, S. C., Porte, D. Jr, Seeley, R. J. & Baskin, D. G. Central nervous system control of food intake. *Nature* **404**, 661–671 (2000).
5. Garofalo, R. S. Genetic analysis of insulin signaling in *Drosophila*. *Trends Endocrinol. Metab.* **13**, 156–162 (2002).

6. Kimura, K. D., Tissenbaum, H. A., Liu, Y. & Ruvkun, G. *daf-2*, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. *Science* **277**, 942–946 (1997).
7. Doyon, C., Drouin, G., Trudeau, V. L. & Moon, T. W. Molecular evolution of leptin. *Gen. Comp. Endocrinol.* **124**, 188–198 (2001).
8. Zhang, Y. *et al.* Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425–432 (1994).
9. Farooqi, I. S. *et al.* Effects of recombinant leptin therapy in a child with congenital leptin deficiency. *N. Engl. J. Med.* **341**, 879–884 (1999).
10. Heymsfield, S. B. *et al.* Recombinant leptin for weight loss in obese and lean adults: a randomized, controlled, dose-escalation trial. *J. Am. Med. Assoc.* **282**, 1568–1575 (1999).
11. De Souza, C. T. *et al.* Consumption of a fat-rich diet activates a proinflammatory response and induces insulin resistance in the hypothalamus. *Endocrinology* **146**, 4192–4199 (2005).
12. Munzberg, H., Flier, J. S. & Bjorbaek, C. Region-specific leptin resistance within the hypothalamus of diet-induced obese mice. *Endocrinology* **145**, 4880–4889 (2004).
13. Bruning, J. C. *et al.* Role of brain insulin receptor in control of body weight and reproduction. *Science* **289**, 2122–2125 (2000).
14. Cohen, P. *et al.* Selective deletion of leptin receptor in neurons leads to obesity. *J. Clin. Invest.* **108**, 1113–1121 (2001).
15. Batterham, R. L. *et al.* Gut hormone PYY_{3–36} physiologically inhibits food intake. *Nature* **418**, 650–654 (2002).
16. Cummings, D. E. *et al.* A preprandial rise in plasma ghrelin levels suggests a role in meal initiation in humans. *Diabetes* **50**, 1714–1719 (2001).
17. Tschop, M., Smiley, D. L. & Heiman, M. L. Ghrelin induces adiposity in rodents. *Nature* **407**, 908–913 (2000).
18. Nakazato, M. *et al.* A role for ghrelin in the central regulation of feeding. *Nature* **409**, 194–198 (2001).
19. Wren, A. M. *et al.* Ghrelin enhances appetite and increases food intake in humans. *J. Clin. Endocrinol. Metab.* **86**, 5992 (2001).
20. Cummings, D. E. *et al.* Plasma ghrelin levels after diet-induced weight loss or gastric bypass surgery. *N. Engl. J. Med.* **346**, 1623–1630 (2002).
21. Di Marzo, V. *et al.* Leptin-regulated endocannabinoids are involved in maintaining food intake. *Nature* **410**, 822–825 (2001).
22. Leibowitz, S. F. & Alexander, J. T. Hypothalamic serotonin in control of eating behaviour, meal size, and body weight. *Biol. Psychiatry* **44**, 851–864 (1998).
23. Leibowitz, S. F., Roosin, P. & Rosenn, M. Chronic norepinephrine injection into the hypothalamic paraventricular nucleus produces hyperphagia and increased body weight in the rat. *Pharmacol. Biochem. Behav.* **21**, 801–808 (1984).
24. Obici, S. *et al.* Central administration of oleic acid inhibits glucose production and food intake. *Diabetes* **51**, 271–275 (2002).
25. Loftus, T. M. *et al.* Reduced food intake and body weight in mice treated with fatty acid synthase inhibitors. *Science* **288**, 2379–2381 (2000).
26. He, W., Lam, T. K., Obici, S. & Rossetti, L. Molecular disruption of hypothalamic nutrient sensing induces obesity. *Nature Neurosci.* **9**, 227–233 (2006).
27. Minokoshi, Y. *et al.* AMP-kinase regulates food intake by responding to hormonal and nutrient signals in the hypothalamus. *Nature* **428**, 569–574 (2004).
28. Andersson, U. *et al.* AMP-activated protein kinase plays a role in the control of food intake. *J. Biol. Chem.* **279**, 12005–12008 (2004).
29. Cota, D. *et al.* Hypothalamic mTOR regulates food intake. *Science* **312**, 927–930 (2006).
30. Strubbe, J. H. & Woods, S. C. The timing of meals. *Psychol. Rev.* **111**, 128–141 (2004).
31. Gibbs, J., Young, R. C. & Smith, G. P. Cholecystokinin decreases food intake in rats. *J. Comp. Physiol. Psychol.* **84**, 488–495 (1973).
32. Emond, M., Schwartz, G. J., Ladenheim, E. E. & Moran, T. H. Central leptin modulates behavioural and neural responsivity to CCK. *Am. J. Physiol.* **276**, R1545–R1549 (1999).
33. Morton, G. J. *et al.* Leptin action in the forebrain regulates the hindbrain response to satiety signals. *J. Clin. Invest.* **115**, 703–710 (2005).
34. Grill, H. J. *et al.* Evidence that the caudal brainstem is a target for the inhibitory effect of leptin on food intake. *Endocrinology* **143**, 239–246 (2002).
35. Elmquist, J. K., Bjorbaek, C., Ahima, R. S., Flier, J. S. & Saper, C. B. Distributions of leptin receptor mRNA isoforms in the rat brain. *J. Comp. Neurol.* **395**, 535–547 (1998).
36. Blevins, J. E., Schwartz, M. W. & Baskin, D. G. Evidence that paraventricular nucleus oxytocin neurons link hypothalamic leptin action to caudal brainstem nuclei controlling meal size. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **287**, R87–R96 (2004).
37. Kelley, A. E., Baldo, B. A., Pratt, W. E. & Will, M. J. Corticostriatal-hypothalamic circuitry and food motivation: integration of energy, action and reward. *Physiol. Behav.* **86**, 773–795 (2005).
38. Rolls, E. T. Taste, olfactory, and food texture processing in the brain, and the control of food intake. *Physiol. Behav.* **85**, 45–56 (2005).
39. Kringelbach, M. L., O'Doherty, J., Rolls, E. T. & Andrews, C. Activation of the human orbitofrontal cortex to a liquid food stimulus is correlated with its subjective pleasantness. *Cereb. Cortex* **13**, 1064–1071 (2003).
40. Kelley, A. E. & Berridge, K. C. The neuroscience of natural rewards: relevance to addictive drugs. *J. Neurosci.* **22**, 3306–3311 (2002).

41. Stuber, G. D., Evans, S. B., Higgins, M. S., Pu, Y. & Figlewicz, D. P. Food restriction modulates amphetamine-conditioned place preference and nucleus accumbens dopamine release in the rat. *Synapse* **46**, 83–90 (2002).
42. Carroll, M. E., France, C. P. & Meisch, R. A. Food deprivation increases oral and intravenous drug intake in rats. *Science* **205**, 319–321 (1979).
43. Fulton, S., Woodside, B. & Shizgal, P. Modulation of brain reward circuitry by leptin. *Science* **287**, 125–128 (2000).
44. Figlewicz, D. P. *et al.* Intraventricular insulin and leptin reverse place preference conditioned with high-fat diet in rats. *Behav. Neurosci.* **118**, 479–487 (2004).
45. Figlewicz, D. P. Adiposity signals and food reward: expanding the CNS roles of insulin and leptin. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **284**, R882–R892 (2003).
46. Flier, J. S. Obesity wars: molecular progress confronts an expanding epidemic. *Cell* **116**, 337–350 (2004).
47. Dhillon, H. *et al.* Leptin directly activates SF1 neurons in the VMH, and this action by leptin is required for normal body-weight homeostasis. *Neuron* **49**, 191–203 (2006).
48. Fan, W., Boston, B. A., Kesterson, R. A., Hraby, V. J. & Cone, R. D. Role of melanocortinergic neurons in feeding and the agouti obesity syndrome. *Nature* **385**, 165–168 (1997).
49. Cowley, M. A. *et al.* Leptin activates anorexigenic POMC neurons through a neural network in the arcuate nucleus. *Nature* **411**, 480–484 (2001).
50. Thornton, J. E., Cheung, C. C., Clifton, D. K. & Steiner, R. A. Regulation of hypothalamic proopiomelanocortin mRNA by leptin in *ob/ob* mice. *Endocrinology* **138**, 5063–5066 (1997).
51. Shutter, J. R. *et al.* Hypothalamic expression of ART, a novel gene related to agouti, is up-regulated in obese and diabetic mutant mice. *Genes Dev.* **11**, 593–602 (1997).
52. Yaswen, L., Diehl, N., Brennan, M. B. & Hochgeschwender, U. Obesity in the mouse model of pro-opiomelanocortin deficiency responds to peripheral melanocortin. *Nature Med.* **5**, 1066–1070 (1999).
53. Seeley, R. J. *et al.* Melanocortin receptors in leptin effects. *Nature* **390**, 349 (1997).
54. Erickson, J., Clegg, K. & Palmiter, R. Sensitivity to leptin and susceptibility to seizures of mice lacking neuropeptide Y. *Nature* **381**, 415–418 (1996).
55. Qian, S. *et al.* Neither agouti-related protein nor neuropeptide Y is critically required for the regulation of energy homeostasis in mice. *Mol. Cell. Biol.* **22**, 5027–5035 (2002).
56. Luquet, S., Perez, F. A., Hnasko, T. S. & Palmiter, R. D. NPY/AgRP neurons are essential for feeding in adult mice but can be ablated in neonates. *Science* **310**, 683–685 (2005).
57. Gropp, E. *et al.* Agouti-related peptide-expressing neurons are mandatory for feeding. *Nature Neurosci.* **8**, 1289–1291 (2005).
58. Bewick, G. A. *et al.* Post-embryonic ablation of AgRP neurons in mice leads to a lean, hypophagic phenotype. *FASEB J.* **19**, 1680–1682 (2005).
59. Xu, A. W. *et al.* Effects of hypothalamic neurodegeneration on energy balance. *PLoS Biol.* **3**, e415 (2005).
60. Lambert, P. D. *et al.* Ciliary neurotrophic factor activates leptin-like pathways and reduces body fat, without cachexia or rebound weight gain, even in leptin-resistant obesity. *Proc. Natl Acad. Sci. USA* **98**, 4652–4657 (2001).
61. Kokoeva, M. V., Yin, H. & Flier, J. S. Neurogenesis in the hypothalamus of adult mice: potential role in energy balance. *Science* **310**, 679–683 (2005).
62. Taniguchi, C. M., Emanuelli, B. & Kahn, C. R. Critical nodes in signalling pathways: insights into insulin action. *Nature Rev. Mol. Cell Biol.* **7**, 85–96 (2006).
63. Bjorbaek, C., Uotani, S., da Silva, B. & Flier, J. S. Divergent signaling capacities of the long and short isoforms of the leptin receptor. *J. Biol. Chem.* **272**, 32686–32695 (1997).
64. Xu, A. W. *et al.* PI3K integrates the action of insulin and leptin on hypothalamic neurons. *J. Clin. Invest.* **115**, 951–958 (2005).
65. Niswender, K. D. *et al.* Intracellular signalling. Key enzyme in leptin-induced anorexia. *Nature* **413**, 794–795 (2001).
66. Niswender, K. D. *et al.* Insulin activation of phosphatidylinositol 3-kinase in the hypothalamic arcuate nucleus: a key mediator of insulin-induced anorexia. *Diabetes* **52**, 227–231 (2003).
67. Kitamura, T. *et al.* Forkhead protein FoxO1 mediates AgRP-dependent effects of leptin on food intake. *Nature Med.* **12**, 534–540 (2006).
68. Spanswick, D., Smith, M. A., Mirshamsi, S., Routh, V. H. & Ashford, M. L. Insulin activates ATP-sensitive K⁺ channels in hypothalamic neurons of lean, but not obese rats. *Nature Neurosci.* **3**, 757–758 (2000).
69. Choudhury, A. I. *et al.* The role of insulin receptor substrate 2 in hypothalamic and β cell function. *J. Clin. Invest.* **115**, 940–950 (2005).
70. van den Top, M., Lee, K., Whyment, A. D., Blanks, A. M. & Spanswick, D. Orexin-sensitive NPY/AgRP pacemaker neurons in the hypothalamic arcuate nucleus. *Nature Neurosci.* **7**, 493–494 (2004).
71. Jaworski, J., Spangler, S., Seeburg, D. P., Hoogenraad, C. C. & Sheng, M. Control of dendritic arborization by the phosphoinositide-3'-kinase-Akt-mammalian target of rapamycin pathway. *J. Neurosci.* **25**, 11300–11312 (2005).
72. Balthasar, N. *et al.* Leptin receptor signaling in POMC neurons is required for normal body weight homeostasis. *Neuron* **42**, 983–991 (2004).
73. Pinto, S. *et al.* Rapid rewiring of arcuate nucleus feeding circuits by leptin. *Science* **304**, 110–115 (2004).
74. Sternson, S. M., Shepherd, G. M. & Friedman, J. M. Topographic mapping of VMH $\rightarrow$ arcuate nucleus microcircuits and their reorganization by fasting. *Nature Neurosci.* **8**, 1356–1363 (2005).
75. Berthoud, H. R. Mind versus metabolism in the control of food intake and energy balance. *Physiol. Behav.* **81**, 781–793 (2004).

Acknowledgements This work was supported by NIH grants, the Diabetes Endocrinology Research Center and Clinical Nutrition Research Unit of the University of Washington, and by a grant from the Murdock Charitable Trust. We acknowledge assistance in manuscript preparation provided by C. Balach; discussions with research fellows and faculty at the University of Washington; and wisdom gained from a conference on feeding behaviour held at the Banbury Center, New York, in May 2006.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence should be addressed to M.W.S. (mschwart@u.washington.edu).

ARTICLES

A juvenile early hominin skeleton from Dikika, Ethiopia

Zeresenay Alemseged¹, Fred Spoor², William H. Kimbel³, René Bobe⁴, Denis Geraads⁵, Denné Reed⁶
& Jonathan G. Wynn^{7†}

Understanding changes in ontogenetic development is central to the study of human evolution. With the exception of Neanderthals, the growth patterns of fossil hominins have not been studied comprehensively because the fossil record currently lacks specimens that document both cranial and postcranial development at young ontogenetic stages. Here we describe a well-preserved 3.3-million-year-old juvenile partial skeleton of *Australopithecus afarensis* discovered in the Dikika research area of Ethiopia. The skull of the approximately three-year-old presumed female shows that most features diagnostic of the species are evident even at this early stage of development. The find includes many previously unknown skeletal elements from the Pliocene hominin record, including a hyoid bone that has a typical African ape morphology. The foot and other evidence from the lower limb provide clear evidence for bipedal locomotion, but the gorilla-like scapula and long and curved manual phalanges raise new questions about the importance of arboreal behaviour in the *A. afarensis* locomotor repertoire.

The skull and associated partial skeleton (DIK-1-1) were recovered at Locality DIK-1 of the Dikika research area during the 2000, 2002 and 2003 field seasons by the Dikika Research Project (DRP)¹ led by Z.A. It derives from sediments of the Sidi Hakoma Member of the Hadar Formation, which span the age of 3.31 to 3.35 million years on the basis of stratigraphic scaling and known chronostratigraphy². Sandstones that yielded the specimen were deposited on a subaerial delta plain. This depositional setting, combined with the remarkable preservation of many articulated faunal remains lacking evidence of preburial weathering, most likely indicates that the juvenile hominin was buried as an intact corpse shortly after death during a major flood event.

The new find represents a young individual. The skull and numerous elements of the articulated axial skeleton were recovered in a block of sandstone matrix (Fig. 1a–f), and additional postcranial parts were found separately (Table 1). As originally recovered, the skull was devoid of covering matrix except for the midface, left temporal bone and the cranial base, but most of the postcranium was covered by matrix (Fig. 1a). Preparation of the main block has been undertaken for the past five years, and this has progressed to a point where much of the specimen's key morphology can be studied. However, fully exposing and isolating the many postcranial elements is a complex task that will take several more years to complete (Fig. 1b, c).

The cranium is intact except for parts of the frontal squama and significant parts of both parietals, which have broken away to reveal the complete natural brain endocast (Fig. 1d). The back of the calvaria is slightly distorted, pushing the nuchal region forward (Fig. 1f). The mandible is complete and is still in articulation with the cranium, but has slipped anteriorly by a few millimetres, which may have played a role in displacing the right zygomaticomaxillary area as well as the nasal and left supraorbital regions in the same

direction (Fig. 1e). All deciduous teeth are preserved, except for the left lower incisor crowns (Fig. 3a). The hyoid bone is preserved beneath the palate (Fig. 3b).

The articulated postcranial elements in the primary sandstone block include both scapulae and clavicles, the cervical, thoracic and the first two lumbar vertebrae, and many ribs. They are displaced from their original anatomical positions, and are compressed superiorly under the cranial base and the palate, making preparation difficult (Fig. 1b, c). Elements of the limbs recovered separately consist of a distal fragment of the right humerus including a tiny unfused medial epicondyle (Fig. 2a), and one ray of the hand containing proximal, intermediate and distal phalanges, along with the proximal phalanx from an adjacent ray (Fig. 3c). The distal femora and proximal tibiae include significant portions of their shafts (of which approximately two-thirds is preserved), and both patellae are present (Fig. 2b, c). The distal ends of the left tibia and fibula, including their unfused epiphyses, are still articulated with the partly matrix-covered foot (Fig. 2d). The latter lacks only its phalanges and the distal-most ends of the metatarsals.

Description and comparison

Age and sex of DIK-1-1. All deciduous teeth were erupted and in occlusion at death. Computed tomography (CT) scanning reveals unerupted, but fully formed, first molar crowns with clear evidence of root development, as well as partial crowns of the second molars (Fig. 3a). Premolar and permanent canine and incisor crowns are also present, with the fourth premolars being least completed. On the basis of this stage of dental development, and using an African ape model³, the age at death of DIK-1-1 is estimated as about three years. Assuming an attribution to *Australopithecus afarensis* (see below), CT-based measurements of the fully formed permanent tooth crowns suggest that the specimen is female (Table 2, Supplementary Note S1).

¹Department of Human Evolution, Max-Planck-Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany. ²Department of Anatomy and Developmental Biology, University College London, Gower Street, London WC1E 6BT, UK. ³Institute of Human Origins, School of Human Evolution and Social Change, PO Box 874101, Arizona State University, Tempe, Arizona 85287-4101, USA. ⁴Department of Anthropology, University of Georgia, Athens, Georgia 30602-1619, USA. ⁵CNRS UPR 2147, 44 rue de l'Amiral Mouchez, F-75014 Paris, France. ⁶Department of Anthropology, University of Texas at Austin, Austin, Texas 78712, USA. ⁷School of Geography and Geosciences, University of St Andrews, St Andrews KY16 9AL, Scotland, UK. †Present address: Department of Geology, University of South Florida, Tampa, Florida 33620, USA.

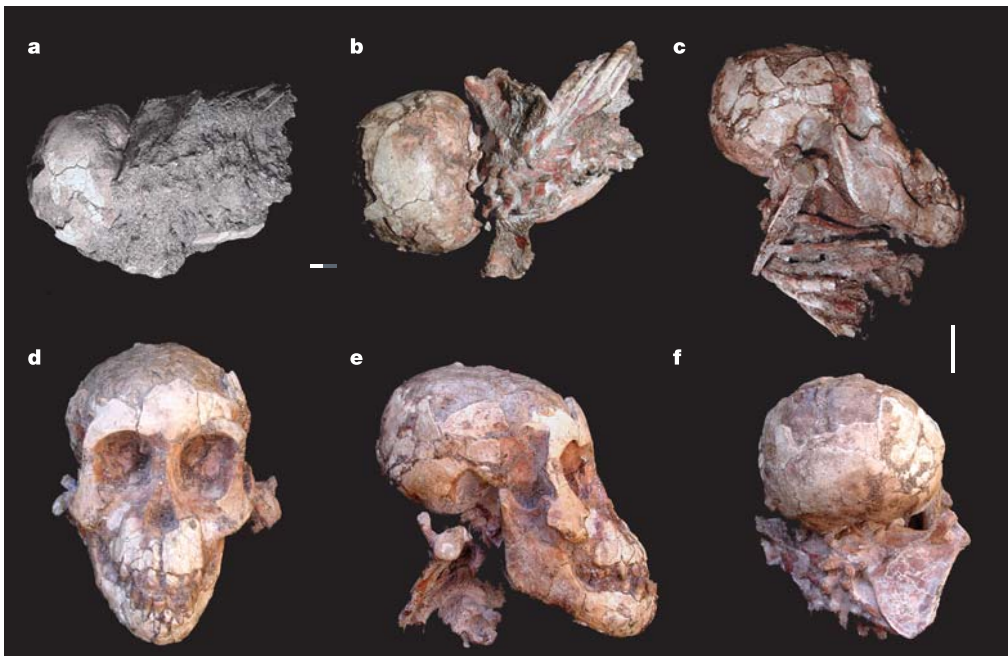


Figure 1 | The DIK-1-1 juvenile skull and partial skeleton. **a, b**, Dorsal and inferior view as discovered (**a**) and after partial preparation (**b**). **c**, Lateral view after partial preparation showing the scapula and many ribs. **d–f**, Anterior (**d**), lateral (**e**) and posterior (**f**) views. Scale bars, 2 cm (**a, b** and **c–f**).

Comparative skull morphology. Size and proportions of the DIK-1-1 face resemble those of the juvenile specimens A.L. 333-105 and Taung, assigned to *A. afarensis* and *A. africanus*, respectively (Table 3). In overall prognathism, it is close to both juvenile and adult specimens of these two species (Table 3, angle sellion–prosthion to alveolar margin). In lateral view, the facial profile forms a relatively straight line from sellion to nasospinale to prosthion (Fig. 1e), whereas in adult *A. afarensis* (for example, A.L. 417-1, 444-2) the midfacial outline is distinctly more vertical than the more prognathic subnasal clivus (Table 3, angle sellion–nasospinale–prosthion). In contrast, older *A. africanus* individuals (Sts 5, 52, 71) have a straighter facial profile than the Taung juvenile. It is not clear if this represents different patterns of facial maturation, as all the variation subsumed by these australopith specimens can be accommodated within the range of intraspecific variation of African apes (Supplementary Note S2).

In the detailed morphology of the face, DIK-1-1 resembles *A. afarensis*⁴ and differs from *A. africanus*, including Taung⁵. Its nasoalveolar clivus is biconvex, whereas it is flatter in *A. africanus*, particularly mediolaterally. The nasal aperture is narrow, as in other juvenile *A. afarensis* such as A.L. 333-86, particularly compared with its height⁶. The canine juga in DIK-1-1 are placed more laterally than in *A. africanus* and they are topographically independent of the sharp-rimmed nasal aperture, as in A.L. 333-105. In contrast to

Taung, no incipient anterior pillars are present and the frontal process of the maxilla is flat lateral to the aperture⁷. The nasal bones of DIK-1-1 are tall, narrow and hourglass-shaped, with nasion positioned well above the nasomaxillary and frontomaxillary sutures. This morphology is seen in A.L. 333-105 and in many apes⁶, but differs from the shorter and broader nasal bones of *A. africanus*. DIK-1-1 has only a slightly convex glabellar region, unlike the more prominent development in Taung. Most of the basicranium is still covered with matrix and displaced axial elements, but the exposed nuchal plane and CT scans suggest that the foramen magnum is located more anteriorly than in apes of comparable dental age^{8,9}.

Preliminary volume measurements of the preserved endocranial of DIK-1-1 from CT scans yield a value of 235 cm³. However, this endocranial volume (EV) underestimates the true value because of the minor deformation of the occipital region, and a few areas of the cranial base where the bone–matrix interface is unclear. To provide an alternative estimate, we calculated the correlation between the EV and the combined endocranial breadth and midsagittal arc for an ontogenetic series of *Pan troglodytes* and *Gorilla gorilla*. Using the regression equations, the EV of DIK-1-1 was estimated as 275 to 330 cm³ (Supplementary Note S3). This is not unlike the volume evident in *P. troglodytes* of a comparable dental age of three years (Supplementary Note S4a). DIK-1-1 would have completed between ~65 and 88 per cent of an average EV of 375 to 425 cm³ estimated for

Table 1 | Dikika fossil hominin specimen discoveries

Specimen number	Collection date	Element	Found by
DIK-1-1a	10 December 2000	Skull and partial skeleton	T. Gebreselassie
DIK-1-1b	30 December 2000	Frontal fragment	R. Abe
DIK-1-1c	30 December 2000	Left scapula fragment plus ribs	Z.A.
DIK-1-1d	19 January 2002	Manual phalanges	Z.A.
DIK-1-1e	25 January 2002	Left proximal tibia	Z.A.
DIK-1-1f	21 January 2002	Left foot	D.G.
DIK-1-1 g	26 January 2003	Right distal femur, patella and proximal tibia	D.R.
DIK-1-1i	4 February 2003	Right humerus	Z.A.
DIK-1-1j	8 February 2003	Left distal femur and patella	A. Ahmed
DIK-1-1 h	17 November 2003	Left tibia fragment	W. Obsie
DIK1-1k	3 December 2003	Left femur fragment	W. Obsie
–	–	Many rib fragments	Team



Figure 2 | The DIK-1-1 juvenile postcrania. **a**, Right distal humerus. **b**, Left distal femur and proximal tibia. **c**, Right distal femur and proximal tibia in a flexed position, connected by matrix, with the tibial and femoral diaphyses pointing upwards. **d**, Left foot and its outline including metatarsals (mt), distal tibia (ti), distal fibula (fi), talus (ta), calcaneus (ca) and tarsals (ts). Scale bars, 1 cm for the foot (**d**) and 2 cm for the limb bones (**a–c**).

adult female *A. afarensis*¹⁰. This proportional endocranial volume, and that of A.L. 333-105, overlaps with the range of variation of both modern humans and African apes (Supplementary Note S4b).

The current position of the hyoid bone beneath the palate precludes a comprehensive analysis of its morphology (Fig. 3b), but some diagnostic features can be observed and measured. It is most similar to that of juvenile African apes, and unlike that of modern humans^{11,12}. The exposed greater horn is slender, and the body is expanded anteriorly, forming a bulla that is deep relative to its breadth and height (Fig. 4a, Supplementary Note S5).

The mandible of DIK-1-1 closely resembles juvenile *A. afarensis* specimens, especially A.L. 333-43 (ref. 13). As in adult Hadar specimens of *A. afarensis*, the external symphyseal contour is bulbous and fairly vertical, distinguishing it from Laetoli juvenile and adult conspecifics, which resemble the more retreating, posteroinferiorly inclined condition seen in *A. anamensis*^{14,15}. As in A.L. 333-43 and A.L. 333n-1, a single anterosuperiorly opening mental foramen is located just below midcorpus at the level of the first deciduous molar¹⁶.

The deciduous upper central incisors are larger than the lateral ones (Table 2), and their overall morphology is similar to that of juvenile apes. In labial view, they are fan-shaped, rather than cylindrical as in Taung, although damage has altered the original

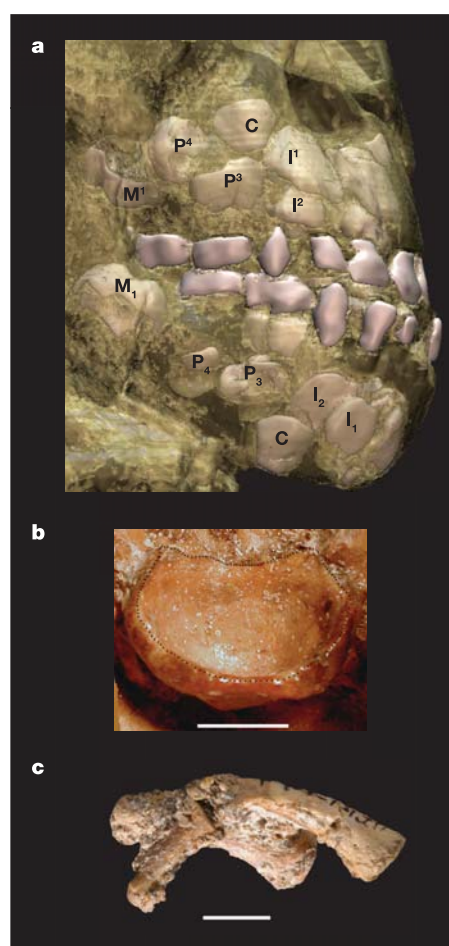


Figure 3 | The DIK-1-1 dentition, hyoid bone and manual phalanges.

a, Three-dimensional reconstruction of CT images of the maxilla and mandible in occlusion showing the deciduous and permanent dentition. Bony elements of the skull have been made translucent to make the dentition visible, and only permanent teeth of the right side are labelled.

b, Inferoposterior view of the body of the hyoid bone sitting on the palate with the anterior bulla facing inferiorly. Dark dots indicate the rim of the body. **c**, Hand phalanges: one ray of the hand containing the proximal and intermediate phalanx is on top; beneath this is the proximal phalanx from an adjacent ray. The three phalanges are connected by matrix. Scale bars, 5 mm (**b**, **c**).

shape of the latter. Crowns of the lower incisors are small and similar in size. The deciduous upper canines have diamond-shaped crowns in labial view, similar to Hadar specimen A.L. 333-99 (ref. 17). In labial view, the lower canine crown is high, pointed and mesiodistally convex; it projects above the occlusal level of the incisors and deciduous molars. The protoconid of the dm1 (first deciduous molar) is larger than the hypoconid, as in A.L. 333-43; the buccal face slopes towards the protoconid tip and is ringed by a basal bulge¹⁸. These features distinguish the DIK-1-1 dm1 from those of great apes, which are buccolingually compressed and almost unicuspid, with the protoconid dominating the occlusal view. As for Hadar specimens, the DIK-1-1 deciduous canines and molars do not show the more ape-like morphology encountered in *A. anamensis* and *A. ramidus*^{19,20}. On the basis of the many craniodental similarities to juvenile and adult Hadar specimens, we attribute DIK-1-1 to *A. afarensis*.

Postcranium of DIK-1-1. Most bipedal features seen in *A. afarensis* specimens are observed on the lower limb and foot of DIK-1-1 (ref. 21). Overall, the tibiae—with their transversely expanded shaft beneath the tibial plateau—are similar to that of the juvenile A.L. 333-39 (ref. 22), but have a sharper anterior border also shown by modern

Table 2 | Dental measurements for DIK-1-1

	MD	BL and LL	CH (buccal)
Upper permanent			
I ¹	10.2	>7.7	>11.5
I ²	7.0	>5.3	>7.3
C	9.5	>7.5	>9.2
P ³	8.5	>11.2	-
P ⁴	≥8.2	>10.2	-
M ¹	11.5	12.4*/12.5†	-
Lower permanent			
I ₁	6.5	>4.7	>10
I ₂	7.2	>5.7	>9.3
C	8.3	>6	>9.5
P ₃	8.9*/9.2†	10.3*/9.8†	-
M ₁	12.3	11.7	-

Values preceded by ">" represent a minimum size; crown possibly not fully formed locally. MD, mesiodistal; BL, buccolingual; LL, labiolingual; CH, crown height.

*Method after ref. 18.

†Method after ref. 36.

humans (Fig. 2b). As in modern humans, the tibialis anterior muscle originated anterior to the interosseous ridge and occupied the lateral side of the tibial shaft, extending the sharp anterior border. The tibialis posterior muscle occupied the lateral aspect of the posterior part of the shaft. Similar to humans, the lateral upper part of the shaft is rather concave, particularly just below the condyles, and becomes convex more distally. Also, the cross-section of the shaft changes from triangular proximally to oval at its distal-most preserved part. The bicondylar angle and symmetrical condyles are additional features previously known in *A. afarensis*. The medial and lateral borders of the talar trochlea are similar in curvature and height (Fig. 2d), and the partly exposed medial side of the talar body is vertical in surface orientation (Fig. 2d). The calcaneus of DIK-1-1 is robust, as in humans, and its distal part is mediolaterally wider in relation to its dorsoplantar dimension compared with that of *Pan*.

The shape of the scapula resembles the scapulae of juvenile and adult gorillas (Fig. 5, Supplementary Note S6). In contrast, modern humans at a similar age have a wider infraspinous fossa and a more laterally facing glenoid fossa, with a correspondingly horizontal spine orientation, whereas chimpanzees tend to have a narrower infraspinous fossa and a more superiorly facing glenoid fossa with a corresponding spine orientation. DIK-1-1 does contrast with gorillas in its narrower supraspinous fossa and less inclined spine, and in these features it is intermediate between African apes and humans. Nevertheless, comparing supraspinous and infraspinous fossa breadths still groups DIK-1-1 more closely with gorilla than with modern humans (Supplementary Note S6e, h). These affinities are also shown in a principal components analysis of 11 linear morphometric measurements (Fig. 4b, Supplementary Note S6i). Finally, the distal humerus preserves a deep and wide olecranon fossa and the

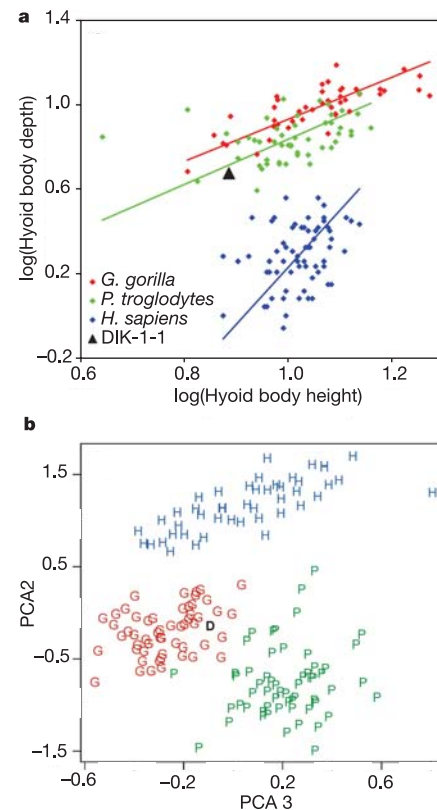


Figure 4 | The DIK-1-1 hyoid and scapula morphology. **a**, Bivariate plot of the posterior depth of the hyoid body against its height (in mm). The values are logged because the relationship is allometric in modern humans (Supplementary Note S5c). **b**, Principal components analysis (PCA) of 11 linear morphometric measurements of the scapula of DIK-1-1 (D), *G. gorilla* (G), *P. troglodytes* (P) and *Homo sapiens* (H). Comparative samples include juveniles and adults (Supplementary Note S6i).

manual phalanges are curved and long, as noted previously for *A. afarensis* (Fig. 3c).

Discussion

The DIK-1-1 skeleton is unmatched in the early hominin fossil record not only because of its completeness, but because it reveals previously unknown morphology and provides a comprehensive cranial and postcranial record of a juvenile australopith. As it concerns a single individual, rather than isolated finds of variable age assumed to represent a single species, it will now be possible to

Table 3 | Hominin cranial measurements

	<i>A. afarensis</i>			<i>A. africanus</i>	
	DIK-1-1a	A.L. 333-105	A.L. 417-1*, 444-2	Taung	Sts 5, 52, 71
Linear dimension (mm)					
Glabella-opisthocranium	>98	NP	-	~127	-
Prosthion-opisthocranium	>133	NP	-	~144	-
Maximum vault breadth	90	NP	-	~95	-
Nasion-nasospinale (na-ns)	32	37	-	33	-
Nasospinale-prosthion (ns-pr)	17	17	-	18	-
Bimaxillary breadth (zm-zm)	60	65	-	65	-
(na-ns/zm-zm) × 100	53	57	-	51	-
(ns-pr/zm-zm) × 100	28	26	-	28	-
Angle (degrees)					
Sellion-prosthion to post-canine alveolar margin	62	59	55-58	64	60-61
Sellion-nasospinale-prosthion	155	149	144-147	140	148-149

Values preceded by ">" represent minimum values for DIK-1-1 because of occipital deformation. Values preceded by "~" are estimates based on the endocast³⁷. NP, not preserved.

*Sellion of A.L. 417-1 estimated.

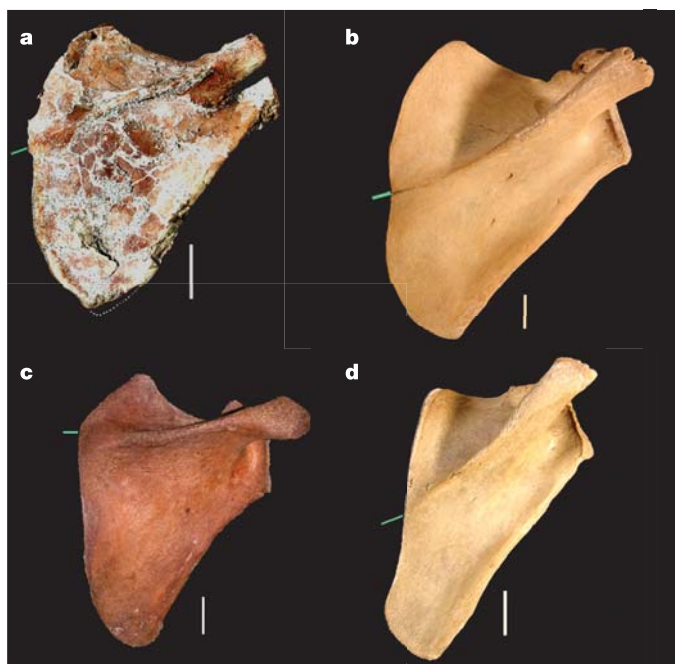


Figure 5 | Right scapulae, dorsal view. **a**, DIK-1-1 (*A. afarensis*). Dotted line around the inferior angle estimates the shape of the broken piece in that region, on the basis of the preserved part on the left scapula of the same individual. **b**, *G. gorilla*. **c**, *H. sapiens*. **d**, *P. troglodytes*. In each, the short green line shows the point where the spine meets the medial border of the scapula. All are juveniles of about the same dental age and oriented with the line connecting the superior and inferior angles perpendicular to the horizontal. Scale bars, 1 cm (**a–d**).

study early hominin ontogenetic development by integrating directly associated evidence from the dentition, brain and postcranium. Its attribution to *A. afarensis*, and the adult context in which it thus should be considered, is unambiguous because the diagnostic facial morphology of this species is evident even at this juvenile stage. Substantial change in facial shape is mainly anticipated in the subnasal region, probably in association with the eruption of the anterior teeth.

Brain growth in DIK-1-1—expressed as the percentage of mean adult EV completed—is at a developmental stage when the patterns of African apes and modern humans overlap substantially and, consequently, the fossil cannot be grouped specifically with either (Supplementary Note S4b). Nevertheless, it is intriguing that in this proportional EV, both DIK-1-1 and A.L. 333-105 fall below the average rate of African apes, and are broadly more in line with the average rate in modern humans. In modern humans, the lower proportional EV rates are due to their large adult EV. However, in *A. afarensis*, with its adult EV values within the range of African apes, the lower proportional EV would have to imply slower absolute brain growth. The possibility of this phenomenon, known in platyrrhines²³, warrants further investigation, but resolution may only come from the discovery of younger individuals, representing a stage when apes and humans show much less overlap in their patterns of brain development.

The hyoid of DIK-1-1 is only the second example in the hominin fossil record¹², and this element was previously unknown for any species earlier than Neanderthals. Its similarities with *Pan* and *Gorilla* hyoids suggest that the bulla-shaped body is the primitive condition for African apes and humans, rather than the more shallow, bar-like body shown by modern humans and *Pongo*. The bulla-shaped body almost certainly reflects the presence of laryngeal air sacs characteristic of African apes²⁴. However, the function of these structures is not well understood²⁵.

The DIK-1-1 skeleton confirms the functional dichotomy of the body plan of *A. afarensis*: a more derived lower body adapted for bipedal locomotion, combined with an upper body that is, in many respects, ape-like. The functional interpretation of these features is highly debated, with some arguing that the upper limb features are non-functional retentions from a common ancestor only^{26–28}, whereas others propose that they were preserved because *A. afarensis* maintained, to some degree, an arboreal component in its locomotor repertoire^{29–32}. Now that the scapula of this species can be examined in full for the first time, it is unexpected to find the strongest similarities with *Gorilla*, an animal in which weight-bearing and terrestrial knuckle-walking predominately characterize locomotor use of the forelimbs³³. Problematic in the interpretation of these findings is that the diversity of scapula architecture among hominoid species is poorly understood from a functional perspective.

The superiorly facing glenoid fossa, noticed previously³¹, provides the most tantalizing suggestion that the structure and function of the upper body in *A. afarensis* was different from that of modern humans. It could indicate a superiorly positioned shoulder girdle, and possibly the absence of effective decoupling of head and trunk movements, typical of modern humans and their capability for endurance running³⁴. The preserved clavicles and full cervical vertebral column of DIK-1-1 will bring new insights in this respect, pending a particularly difficult process of preparation to isolate these elements. One further clue in this context is that the semicircular system in DIK-1-1 is similar to that of African apes and *A. africanus* (Supplementary Note 7), and this has been associated with limited head decoupling and absence of fast and agile bipedal gaits³⁵. If functionally relevant, the glenoid fossa orientation in DIK-1-1 could also point to frequent use of the arms above the head³¹, and the activity with which this would be most consistent is climbing. The apparent tendency towards a reduced attachment area of the supraspinous muscle could be seen as evidence for a diminishing role of the arms in such locomotor postures. The scapula morphology, together with forelimb features such as the long and curved manual phalanges of DIK-1-1, will raise new questions about the importance of arboreal behaviour in *A. afarensis*. After years of discussion, new data, and in particular the ability to reconstruct much of an entire body of a three-year-old *A. afarensis* child will bring important new evidence to bear on this debate.

Received 22 April; accepted 6 July 2006.

1. Alemseged, Z. *et al.* A new hominin from the Basal Member of the Hadar Formation, Dikika, and its geological context. *J. Hum. Evol.* **49**, 499–514 (2005).
2. Wynn, J. G. *et al.* Geological and palaeontological context of a Pliocene juvenile hominin at Dikika, Ethiopia. *Nature* doi:10.1038/nature05048 (this issue).
3. Dean, C. & Wood, B. A. Digital radiographic atlas of the great apes skull and dentition. In *Digital Archives of Human Paleobiology* (eds Bondioli, L. & Machiarelli, R.) CD-ROM Vol. 3 (Museo Nazionale Preistorico Etnografico “L. Pigorini”, Rome, 2003).
4. Kimbel, W. H., Rak, Y. & Johanson, D. C. *The Skull of Australopithecus afarensis* (Oxford Univ. Press, New York, 2004).
5. Dart, R. A. *Australopithecus africanus*: The man-ape of South Africa. *Nature* **115**, 195–199 (1925).
6. Kimbel, W. H., Johanson, D. C. & Coppens, Y. Pliocene hominid cranial remains from the Hadar Formation, Ethiopia. *Am. J. Phys. Anthropol.* **57**, 453–499 (1982).
7. Rak, Y. *The Australopithecine Face* (Academic, New York, 1983).
8. Schultz, A. H. The position of the occipital condyles and of the face relative to the skull base in primates. *Am. J. Phys. Anthropol.* **13**, 97–120 (1955).
9. Luboga, S. A. & Wood, B. A. Position and orientation of the foramen magnum in higher primates. *Am. J. Phys. Anthropol.* **81**, 67–76 (1990).
10. Holloway, R. L., Broadfield, D. C. & Yuan, M. S. *Brain Endocasts—The Paleoneurological Evidence. The Human Fossil Record Vol. 3* (Wiley, New Jersey, 2004).
11. Clegg, M. *The Comparative Anatomy and Evolution of the Human Tract*. PhD thesis, Univ. College London (2001).
12. Arensburg, B. *et al.* A Middle Palaeolithic human hyoid bone. *Nature* **338**, 758–760 (1989).
13. White, T. D. & Johanson, D. C. Pliocene hominid mandibles from the Hadar Formation, Ethiopia: 1974–1977 collections. *Am. J. Phys. Anthropol.* **57**, 501–544 (1982).

14. Ward, C. V., Leakey, M. G. & Walker, A. Morphology of *Australopithecus anamensis* from Kanapoi and Allia Bay, Kenya. *J. Hum. Evol.* **41**, 255–368 (2001).
15. Kimbel, W. H. et al. Was *Australopithecus anamensis* ancestral to *A. afarensis*? A case of anagenesis in the Pliocene hominin fossil record. *J. Hum. Evol.* **51**, 134–152 (2006).
16. Johanson, D. C. The most primitive *Australopithecus*. in *Hominid Evolution: Past, Present and Future* (ed. Tobias, P. V.) 203–212 (Wiley, New York, 1985).
17. Johanson, D. C., White, T. D. & Coppens, Y. Dental remains from the Hadar Formation, Ethiopia: 1974–1977 collections. *Am. J. Phys. Anthropol.* **57**, 545–603 (1982).
18. White, T. D. New fossil hominids from Laetolil, Tanzania. *Am. J. Phys. Anthropol.* **46**, 197–230 (1977).
19. White, T. D., Suwa, G. & Asfaw, B. *Australopithecus ramidus*, a new species of early hominid from Aramis, Ethiopia. *Nature* **371**, 306–312 (1994).
20. Leakey, M. G., Feibel, C. S., McDougall, I., Ward, C. & Walker, A. New specimens and confirmation of an early age for *Australopithecus anamensis*. *Nature* **393**, 62–66 (1998).
21. Aiello, A. & Dean, C. *An Introduction to Human Evolutionary Anatomy* (Academic, London, 1990).
22. Lovejoy, O. C., Johanson, D. C. & Coppens, Y. Hominid lower limb bones recovered from the Hadar Formation: 1974–1977 collections. *Am. J. Phys. Anthropol.* **57**, 679–700 (1982).
23. Leigh, S. R. Brain growth, life history, and cognition in primate and human evolution. *Am. J. Primatol.* **62**, 139–164 (2004).
24. Avril, C. Kehlkopf und Kehlsack des Schimpansen (*Pan troglodytes*). *Gegenbaurs Morphol. Jahrb.* **105**, 74–129 (1963).
25. Hewitt, G., MacLarnon, A. & Jones, K. E. The functions of laryngeal air sacs in primates: a new hypothesis. *Folia Primatol. (Basel)* **73**, 70–94 (2002).
26. Lovejoy, C. O. The origin of man. *Science* **211**, 341–350 (1981).
27. Lovejoy, C. O. Evolution of human walking. *Sci. Am.* **259**, 118–125 (1988).
28. Latimer, B. Locomotor adaptations in *Australopithecus afarensis*: the issue of arboreality. in *Origine de la bipédie chez les Hominidés* (eds Senut, B. & Coppens, Y.) 169–176 (CNRS, Paris, 1991).
29. Senut, B. & Tardieu, C. Functional aspects of Plio-Pleistocene hominid limb bones: implications for taxonomy and phylogeny. in *Ancestors: The Hard Evidence* (ed. Delson, E.) 193–201 (Liss, New York, 1979).
30. Susman, R. L., Stern, J. T. & Jungers, W. L. Arboreality and bipedality in the Hadar hominids. *Folia Primatol. (Basel)* **43**, 113–156 (1984).
31. Stern, J. T. & Susman, R. L. The locomotor anatomy of *Australopithecus afarensis*. *Am. J. Phys. Anthropol.* **60**, 279–317 (1983).
32. Stern, J. T. & Susman, R. L. “Total morphological pattern” versus the “magic trait”: conflicting approaches to the study of early hominid bipedalism. in *Origine de la bipédie chez les Hominidés* (eds Senut, B. & Coppens, Y.) 99–112 (CNRS, Paris, 1991).
33. Richmond, B. G. & Strait, D. S. Evidence that humans evolved from a knuckle-walking ancestor. *Nature* **404**, 382–385 (2000).
34. Bramble, D. M. & Lieberman, D. E. Endurance running and the evolution of *Homo*. *Nature* **432**, 345–352 (2004).
35. Spoor, F., Wood, B. & Zonneveld, F. Implications of early hominid labyrinthine morphology for the evolution of human bipedal locomotion. *Nature* **369**, 645–648 (1994).
36. Wood, B. *Koobi Fora Research Project Vol. 4: Hominid Cranial Remains* (Clarendon, Oxford, 1991).
37. Broom, R. & Schepers, G. W. H. *The South African Fossil Ape-men: The Australopithecinae*. Transvaal Museum Memoir 2 (Transvaal Museum, Pretoria, 1964).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Authority for Research and Conservation of Cultural Heritage (ARCCH) and the National Museum of Ethiopia of the Ministry of Tourism and Culture, and the Afar regional government for permits; A. Dessie, T. Gebreselassie, G. Assefa, A. Kebede, M. Mekonnen, H. Defar, A. Zerihun, A. Takele and the people of the Dikika area for field assistance; A. Admassu for laboratory assistance; the National Geographic Society, the Institute of Human Origins at ASU and the Leakey Foundation for financial and logistical support. For discussions and help, we thank R. Ackermann, M. Clegg, H. Coqueugnot, C. Dean, T. Fitch, J.-J. Hublin, L. Humphrey, P. Jenkins, D. Johanson, K. Kupczik, S. Larson, D. Lieberman, C. Lockwood, R. Martin, E. Mbua, R. McCarthy, K. Mowbray, P. Morris, T. Nishimura, P. O'Higgins, O. Pearson, A. Pommert, Y. Rak, J. Richtsmeier, L. Scheuer, B. Sokhi, B. Viola, C. Ward and T. Weaver. We also thank B. Wood for critical comments and suggestions that have improved our paper greatly.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Z.A. (zeray@eva.mpg.de).

ARTICLES

Crystal structure of an H/ACA box ribonucleoprotein particle

Ling Li¹ & Keqiong Ye¹

H/ACA ribonucleoprotein particles (RNPs) are a family of RNA pseudouridine synthases that specify modification sites through guide RNAs. They also participate in eukaryotic ribosomal RNA processing and are a component of vertebrate telomerases. Here we report the crystal structure, at 2.3 Å resolution, of an entire archaeal H/ACA RNP consisting of proteins Cbf5, Nop10, Gar1 and L7ae, and a single-hairpin H/ACA RNA, revealing a modular organization of the complex. The RNA upper stem is bound to a composite surface formed by L7ae, Nop10 and Cbf5, and the RNA lower stem and ACA signature motif are bound to the PUA domain of Cbf5, thereby positioning middle guide sequences so that they are primed to pair with substrate RNA. Furthermore, Gar1 may regulate substrate loading and release. The structure rationalizes the consensus structure of H/ACA RNAs, suggests a functional role of each protein, and provides a framework for understanding the mechanism of RNA-guided pseudouridylation, as well as various cellular functions of H/ACA RNP.

Pseudouridine, the most commonly modified nucleoside, is post-transcriptionally converted from uridine in a site-specific manner by either stand-alone enzymes that select the modification site via protein recognition, or RNP enzymes that specify target through a bound H/ACA guide RNA that base pairs with the complementary substrate RNA^{1–3}. To guide the formation of numerous pseudouridines in ribosomal and spliceosomal RNAs, there are abundant H/ACA species, with 29 species documented in budding yeast⁴ and ~100 species in humans⁵, forming one of the largest families of non-coding RNAs.

H/ACA RNAs fold into repeats of a consensus hairpin structure comprising a large internal loop and an ACA-motif-harboring 3' tail (Fig. 1a). The internal loop sequence, which comprises the pseudouridylation pocket, can form two short duplexes with substrate RNA and specifies a central, unpaired uridine ~14 nucleotides upstream of the ACA box for modification^{6,7}. Most eukaryotic H/ACA RNAs, existing as small nucleolar RNA (snoRNA) or small Cajal body RNA (scaRNA), have two hairpins, where the 5' hairpin is followed by the H box (ANANNA, where N indicates any base), and the 3' hairpin is followed by the ACA box that is located exactly 3 nucleotides upstream of the 3' end^{8,9}. Archaeal RNAs possess variable hairpins that normally contain an L7ae-binding k-turn motif in the upper stem¹⁰. Most H/ACA RNAs serve as pseudouridylation guides; however, the conserved U17/snoR30 RNA participates in ribosomal RNA processing^{11,12}, and vertebrate telomerase RNAs contain an H/ACA domain essential for *in vivo* telomerase function¹³.

To form a functional complex, each H/ACA RNA associates with a common set of highly conserved proteins^{8,10,14–18}: the pseudouridylation catalyst Cbf5 (dyskerin in humans), the RNA-binding protein L7ae in archaea (Nhp2 in eukaryotes), Nop10 and Gar1. The function of Nop10 and Gar1 remains unclear, although biochemical studies^{19–22} and recent co-crystal structures of Cbf5–Nop10–Gar1 or Cbf5–Nop10 complexes^{23–25} have revealed that Nop10 and Gar1 directly associate with Cbf5. To understand the organization of the complex, the basis of H/ACA RNA consensus structure, and the mechanism of RNA-guided pseudouridylation, we have determined

the crystal structure of an entire archaeal H/ACA RNP and built a model of the substrate complex.

Structural overview

In the 2.3 Å resolution structure (Fig. 1), the 65-nucleotide single-hairpin H/ACA RNA, derived from Afu46 (ref. 10), associates with one copy each of Cbf5, Gar1, Nop10 and L7ae from *Pyrococcus furiosus*, resulting in a 94-kDa complex—RNA comprises one-quarter of the mass of this complex. Such protein-binding stoichiometry probably holds for individual hairpin units in two- and multiple-hairpin H/ACA RNAs.

The protein part resembles an equilateral triangle with sides of 80–90 Å in length, ~20 Å thickness, and with the catalytic domain of Cbf5 (residues 39–292) in the centre, the PUA domain of Cbf5 (residues 1–38, 293–343), L7ae (124 residues) and Gar1 (97 residues) at the three corners, and the smallest protein, Nop10 (60 residues), sandwiched between L7ae and Cbf5. The catalytic domain of Cbf5 can be divided into subdomains D1 and D2 that join through the formation of a continuous β-sheet with the catalytic cleft created between them. Nop10 adopts an extended structure consisting of an amino-terminal zinc ribbon that contacts Cbf5 D2, and a linker region and a carboxy-terminal α-helix that both bind Cbf5 D1. Gar1 forms a six-stranded β-barrel and associates with Cbf5 D2. The structure of Cbf5–Nop10–Gar1 in the full RNP closely resembles that of the previously reported subcomplex²³ with root mean square deviation (r.m.s.d.) of 0.69 Å over 420 aligned Cα atoms.

The RNA hairpin adopts a near-linear structure consisting of the upper helical stem (P2) that contains a k-turn motif, single-stranded 5' and 3' guide sequences, the lower helical stem (P1), and the 3' ACA motif, in agreement with its predicted secondary structure. Nucleotides downstream of the ACA motif are not visible in the crystal, probably due to disorder. Unlike many other structured RNAs, no tertiary interactions occur among these elements. The RNA spans ~80 Å in the long dimension, binds the protein complex on the front face where the active site is located, and fully covers the side spanning the PUA domain and L7ae. The hairpin is fastened to a composite surface formed by L7ae, Nop10 and Cbf5 at the P2 stem,

¹National Institute of Biological Sciences, 7 Science Park Road, Zhongguancun Life Science Park, Beijing 102206, China.

and to the PUA domain at the P1 stem and ACA tail regions, positioning the middle guide sequence loosely over Cbf5 D1 on one side of the active cleft. Such two-end binding creates a 50° bend towards the protein in the helical axis of the P1 and P2 stem and buries a total of 4,680 Å² of solvent-accessible surface.

Cbf5, Nop10 and L7ae directly bind RNA, whereas Gar1 makes no contact with RNA, consistent with its lack of *in vitro* RNA-binding activity¹⁹. Eukaryotic Gar1 may not be a primary RNA assembly factor either, because its core domain shares significant sequence homology with the archaeal protein and its additional glycine-arginine-rich (GAR) domains are functionally dispensable²⁶. The different roles in RNA assembly would explain why depletion of Cbf5, Nop10 and Nhp2, but not Gar1, destabilize H/ACA RNAs in yeast^{14–18}.

PUA recognition of the lower stem and the ACA motif

The lower stem P1 and the ACA motif adopt an 'L-shaped' conformation and bind the PUA domain of Cbf5. The mechanism of RNA recognition differs markedly from the previously characterized PUA recognition mechanism of the acceptor stem and CCA tail of transfer RNA²⁷ (Supplementary Fig. 3). Numerous protein contacts occur at the minor groove via hydrogen bonds with 2'-hydroxyls, phosphate groups and base edges (Fig. 2a). The crystal structure indicates that disruption of the P1 duplex would disrupt RNA binding, and hence lead to inactivation of RNA⁸.

The first and third adenines in the ACA and related H box are universally conserved and critical for RNA stability, localization and function^{8,9}. Both adenines are bound in a sequence-specific manner. The first adenine, A58, in a syn conformation, inserts into an RNA–protein hybrid pocket formed between ribose moieties from the RNA minor groove and protein residues His 268, Gly 269, Ala 270 and Trp 337 (Fig. 2d). The pocket formation is partially induced by RNA binding, as Trp 337 at the bottom of the pocket is otherwise disordered in the structures of free protein^{23,24}. Directional

interactions along the Hoogsteen edge of adenine provide sequence specificity, with N7 of A58 accepting a hydrogen bond from 2'-OH of C56 and N6 of A58 donating a hydrogen bond to His 268 O.

The third adenine, A60, also in a syn conformation, stacks over helix α 8 and lies partially beneath the ribose moiety of the preceding residue C59 (Fig. 2c, d). A60 forms two Watson–Crick-like hydrogen bonds with protein backbone, thereby imposing sequence specificity. Specificity is further determined by hydrogen bonds with 2'-OH of A58 and van der Waals interactions involving C2 that sterically exclude N2 of guanine.

In contrast to burial of adenines, the middle cytosine (C59) adopts a fairly open conformation (Fig. 2c, d), despite the hydrophobic contact with Ile 319 and hydrogen bonding with N of Gly 318. The exposed conformation suggests that specificity at this position is less strict and may be open to modulation by other interactions. In fact, compared with the two adenines, the second base is less conserved as cytosine in the ACA box, with purine preferred in the related H box, and being exclusively guanine in single-hairpin H/ACA RNAs of early diverging eukaryotes *Trypanosoma brucei*²⁸ and *Euglena gracilis*²⁹.

Implications for dyskeratosis congenita

Mutations in dyskerin (human Cbf5) have been linked to the X-chromosome-linked form of dyskeratosis congenita, a rare skin and bone marrow failure syndrome probably related to defects in telomerase and/or ribosome^{30,31}. Most of the dyskeratosis congenita mutations that could be mapped to archaeal Cbf5 occur in the PUA domain^{23–25} (Fig. 2d), with three in the catalytic domain (Supplementary Fig. 4). These mutations in the PUA domain form a spatial cluster and have been suggested to affect RNA association. Notably, our structure shows that they do not overlap with the RNA-binding sites of the P1 stem and the ACA motif, suggesting that they may affect some functions and structures that are present in dyskerin but absent in archaeal Cbf5 (Supplementary Information).

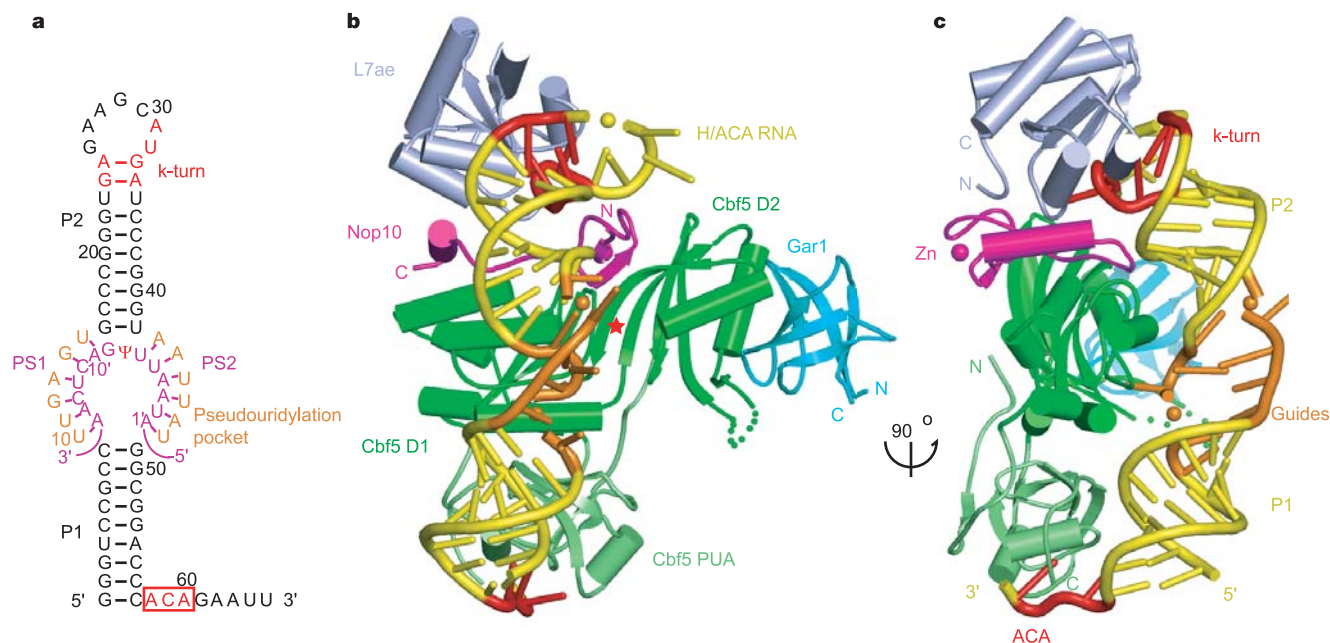


Figure 1 | Structure of H/ACA RNP. **a**, Sequence and secondary structure of the single-hairpin archaeal H/ACA RNA used in this study. Also shown is a bound cognate substrate RNA, which is, however, not present in the structure. Paired regions are named as P1 and P2 for the lower and upper stems, and PS1 and PS2 for duplexes formed between guide sequences (orange) and substrate RNA (purple), respectively. The modification target is shown as product pseudouridine (ψ , red). The terminal k-turn and ACA motif are highlighted in red. Every tenth nucleotide is numbered, with prime

denoting numbering of the substrate RNA. **b**, Ribbon representation of the H/ACA RNP structure in front view. **c**, Side view. Different colours are used for the Cbf5 catalytic domain (dark green), Cbf5 PUA domain (lime green), Nop10 and zinc ion (magenta), L7ae (light blue), Gar1 (cyan), k-turn and ACA motifs (red), guides (orange) and the rest of the RNA (yellow). The same colour coding is used for the other figures. Major secondary structure elements are shown for proteins. The N and C termini are indicated when appropriate. Dots represent disordered chains; star denotes the active site.

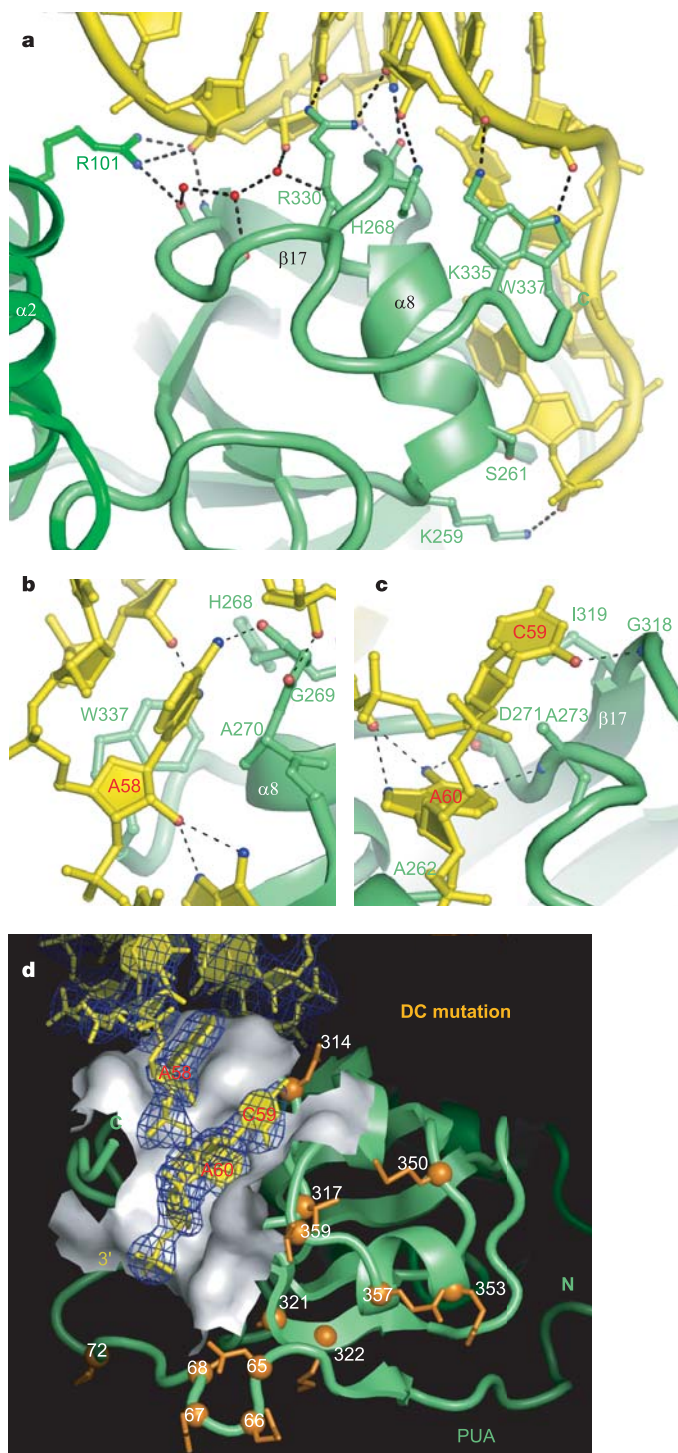


Figure 2 | PUA recognition of the P1 stem and the ACA motif. **a**, Overall view highlighting the intimate protein contacts along the P1 minor groove. **b**, The RNA-protein hybrid binding pocket for A58. **c**, Binding of C59 and A60. Interacting residues are shown in ball-and-stick representation, water as a red sphere, hydrogen bonds as dashed lines, and hydrogen-bonding atoms are coloured as red for oxygen and blue for nitrogen. Interactions are shown in the same manner in Figs 3 and 4. **d**, Dyskeratosis congenita mutations in the PUA domain, indicated by $\text{C}\alpha$ spheres, side chains from this structure and residue numbers in dyskerin. Also shown is the RNA enveloped in the $2f_o - f_c$ electron density map at the 1.5σ level and the partial RNA-binding surface.

Composite binding of the upper stem of H/ACA RNA

The sequential interactions among L7ae, Nop10 and Cbf5 D1 generate a combined surface for binding the P2 stem. The k-turn motif of P2 in our structure is closed by a loop instead of a duplex as in the canonical type¹⁰; nevertheless, the abbreviated k-turn similarly binds L7ae and has the same core structure characterized by tandem sheared GA pairs (G24–A34, G33–A25), a sharp turn in the RNA backbone around U32 and stacking A31 (Fig. 3a; see also Supplementary Information).

Previous biochemical studies showed no interaction of L7ae with the other proteins in the absence of RNA^{19,20}; however, our structure indicates that in the RNA complex L7ae binds at the trough of Nop10 formed between its N-terminal zinc ribbon and C-terminal helix $\alpha 1$ (Fig. 3b). L7ae helix $\alpha 3$ contacts Nop10 $\alpha 1$ at a right angle, which in turn contacts Cbf5 $\alpha 6$ at the opposite face also at a right angle, resulting in a 3-helix sandwich arrangement. The inter-helix interactions between L7ae and Nop10 are mediated by a salt bridge and three aromatic rings (Tyr 41, Tyr 44 and Trp 48) of Nop10, with the latter stacking over a hydrophobic surface on L7ae (Thr 41, Ala 65, His 66, Pro 69 and Leu 70). The loop positioned N-terminally to L7ae helix $\alpha 3$ additionally contacts the linker region of Nop10 through van der Waals and hydrogen bond interactions. Unlike the extensive interface between Nop10 and Cbf5 ($1,490 \text{ \AA}^2$), the interface between Nop10 and L7ae is small ($660 \text{ \AA}^2$), which may account for the lack of stable protein association of L7ae by itself^{19,20}.

In addition to organizing the positioning of L7ae, Nop10 binds the proximal part of P2 with an extremely conserved segment (residues 34–38) that simultaneously makes intimate contact with Cbf5 (Fig. 3d). The rings of Phe 35 and Pro 37 anchor at a hydrophobic patch on helices $\alpha 1$ and $\alpha 6$ of Cbf5, orienting three neighbouring residues (Arg 34, Thr 36 and Glu 38) for interaction with RNA. The long side chain of Arg 34 bridges two phosphate groups across the RNA major groove. Thr 36 and Glu 38, together with Glu 64 of Cbf5, interact with 2'-hydroxyls and bases along the RNA minor groove. Nop10 apparently needs to be in the complex to adopt the conformation for binding RNA. In fact, free Nop10 is largely unfolded²⁴ and displays no RNA-binding activity^{19,20}. The interweaved interactions with Cbf5, L7ae and RNA make Nop10 the central structural organizer at the upper stem region.

Because binding of the P2 stem starts at its distal end, the duplex length of P2 determines the location of its basal end near which substrate RNA is about to associate. To ensure proper positioning of substrate RNA, the P2 length measured between the distal GA pair (A25–G33) and the basal pair (G16–U42) is conserved as 9–10 base pairs (bp) in archaeal H/ACA RNAs¹⁰.

Plastic pseudouridylation pocket

The binding of the upper and lower stem places the internal pseudouridylation pocket near the active cleft of Cbf5. The 5' guide (residues 10–15) is situated away from the Cbf5 surface, whereas the 3' guide (residues 43–48) is proximal to Cbf5 and makes a few intermolecular contacts (Fig. 3c). Notably, protruding U45 stacks over the imidazole ring of His 80 and forms two hydrogen bonds with His 80 backbone atoms. Although binding of U45 seems to be sequence specific, statistics from over 100 H/ACA RNAs reveal no sequence preference at the equivalent position. All pocket nucleotides show high temperature factors and two of them could not be modelled due to missing electron density. Crystal packing interactions stabilize a few of the pocket nucleotides (U10, A13, A47, U48), which may be more mobile when not in the crystal environment. The flexibility of the pocket in the substrate-free structure could potentially facilitate the adoption by guide sequences of new conformations on substrate binding.

Proposed mechanism of RNA-guided pseudouridylation

Substrate binding at the pseudouridylation pocket would generate a

three-helix junction surrounding the target uridine and another unpaired nucleotide⁶ (Fig. 1a). The three helices are P2, PS1 (formed by the 5' guide and the 3' half of substrate RNA) and PS2 (formed by the 3' guide and the 5' half of substrate RNA). We have built a model of the substrate complex (see Methods and Supplementary Fig. 6a), with PS2 and the adjacent target uridine modelled according to the structure of a hairpin substrate RNA bound in the active cleft of TruB³², which is a stand-alone pseudouridine synthase for tRNA Ψ 55.

In the model, two guide strands are moderately displaced from their original position on substrate binding (Supplementary Fig. 6b). The bound substrate RNA, in a 'U' shape, approaches the active cleft in a nearly vertical direction relative to the protein plane (Fig. 4a). The PS2 and P1 stems coaxially stack and form a pseudo-continuous helix that bridges the target uridine and the ACA motif at two ends. Because the ACA motif is bound to the PUA domain, to place the target uridine in the active site properly, the distance between the ACA motif and the target uridine is constrained as ~ 14 nucleotides^{6,7}.

A loop region (140–155) in the Cbf5 D2 subdomain, which is partially disordered in our structure, is expected to bind and stabilize the loaded substrate RNA, like the corresponding region in TruB^{32,33} (Supplementary Fig. 6a). We notice that the loop in the structure of the Cbf5–Nop10–Gar1 complex²³, which contains no RNA, is entirely ordered and coincidentally occupies a position proximal to the modelled substrate RNA; we use the term 'closed-like' state to refer to the loop in the Cbf5–Nop10–Gar1 free complex, which serves as an approximation of the RNA-bound state. The conformation of

that loop may be shaped by contacts with neighbouring crystallographic molecules²³, and the loop in our structure is not involved in crystal packing interactions.

The loop in our structure differs notably from the closed-like state, apart from seven disordered residues (146–152). In both structures, the basal part of the loop similarly binds Gar1 through residue Ile 140, but only in our structure do residues Pro 143, Pro 144 and Leu 145 dock at a hydrophobic patch (Val 23, Val 44 and Leu 26) of Gar1 (Fig. 4b). Notably, these docking interactions, if not disrupted, would prevent the loop from binding substrate RNA because the Pro 144 C α atom needs to travel ~ 15 Å to assume the closed-like state. We propose that the loop adopts two functional states: the closed state where the loop stabilizes loaded substrate, and the 'open' state, observed in our structure, where the loop is unable to contact substrate and thereby facilitates its loading and release. Gar1 may be involved in regulating substrate loading and release by stabilizing the open state of the substrate-binding loop. Consistent with this proposal, H/ACA RNP lacking Gar1 catalyses modification, but at a much slower rate²⁰.

Conclusion

Separating the function of substrate recognition to guide RNA allows H/ACA RNPs to maintain an invariant structural core and specifically target different substrates once programmed with different guide RNAs. To assist the guide RNA, multiple proteins and domains have, during the course of evolution, been incorporated into the catalytic core to form a complex pseudouridine synthase. Our structure reveals the first overall view of how the molecular machine

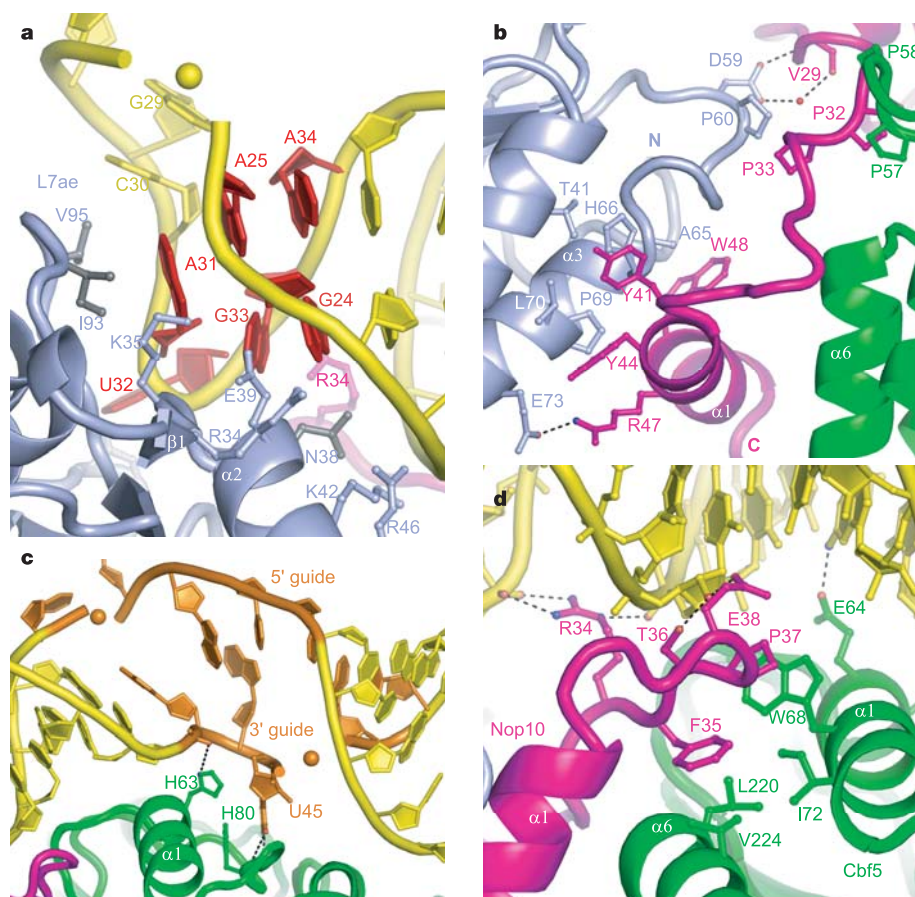


Figure 3 | Recognition of the P2 stem and pseudouridylation pocket. **a**, L7ae and Nop10 binding to the terminal k-turn structure. RNA-binding residues are coloured in light blue for those conserved in both L7ae and Nhp2 and in grey for those that change identity in Nhp2. The k-turn bases

are in red. **b**, The L7ae–Nop10 interface with RNA omitted for clarity. **c**, Flexible pseudouridylation pocket in the substrate-free structure. **d**, Dual binding of the P2 stem by Nop10 and Cbf5.

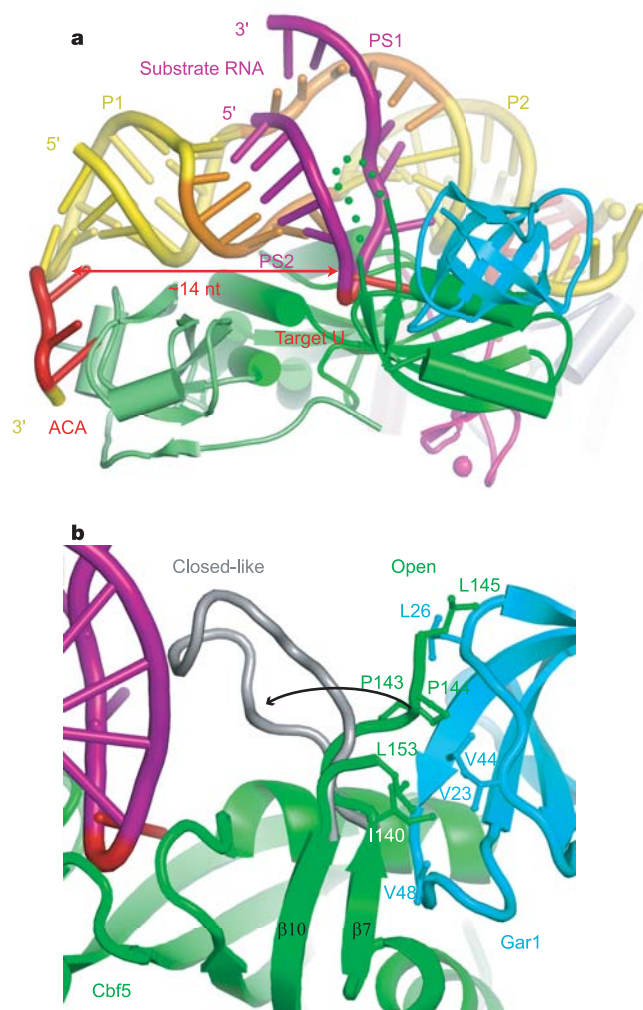


Figure 4 | Proposed mechanism of RNA-guided pseudouridylation. **a**, Structural model of the substrate complex. **b**, Two states of the putative substrate-binding loop. In our structure, the loop adopts an open conformation: it docks against Gar1 and appears to be unable to bind substrate RNA. The loop from the Cbf5–Nop10–Gar1 structure²³ (grey) is shown as an approximation of the RNA-bound state.

is constructed and suggests how it works. Cbf5, Nop10 and L7ae directly coordinate H/ACA RNA in the complex and spatially align guide sequences for binding with substrate, whereas Gar1 may control substrate loading and release. Despite some differences, eukaryotic RNPs probably assemble and function in a similar way to archaeal RNPs (Supplementary Information). This structure provides a new platform from which to understand the various functions of H/ACA RNPs in RNA-guided pseudouridylation, rRNA processing and telomerase activity.

METHODS

A detailed description of protein preparation, structure determination, model building and activity assay is provided in Supplementary Information. The H/ACA complex assembled from full-length recombinant *Pyrococcus furiosus* Cbf5, Nop10, Gar1 and L7ae proteins and *in-vitro*-transcribed guide RNA can direct site-specific pseudouridylation of a cognate substrate RNA (Supplementary Information), as reported previously^{19,20}. Crystallization was done by mixing 1 μ l of 5 mg ml⁻¹ complex in 10 mM MgCl₂, 70 mM EDTA and 5 mM HEPES-Na (pH 7.6) with 1 μ l of 30% methyl-2,4-pentanediol, 35 mM magnesium acetate, 10 mM adenosine-5'-triphosphate (ATP) and 50 mM sodium cacodylate (pH 5.0) and by the hanging-drop vapour diffusion method at 30 °C. The crystal belongs to space group $P2_12_12_1$ ($a = 83.031$ Å, $b = 90.950$ Å and $c = 114.064$ Å), has one complex per asymmetric unit and 45% solvent

content. Diffraction data were collected at Japan SPring-8 BL41XU beamline. The structure was determined by molecular replacement using the structure of the Cbf5–Nop10–Gar1 complex as a search model²³ and refined to a free R -factor of 0.278 at 2.3 Å resolution (Fig. 2d and Supplementary Table 1). The current model includes Cbf5 (residues 11–145 and 153–337), Gar1 (1–74), Nop10 (3–55), L7ae (4–124), RNA (1–13, 15–25, 27–45 and 47–60), 1 zinc ion, 1 ATP and 118 water molecules. ATP was included in crystallization and found to mediate crystal packing. Structures were drawn with PyMol³⁴.

Received 16 May; accepted 9 August 2006.

Published online 30 August 2006.

- Ofengand, J. Ribosomal RNA pseudouridines and pseudouridine synthases. *FEBS Lett.* **514**, 17–25 (2002).
- Kiss, T. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell* **109**, 145–148 (2002).
- Meier, U. T. The many facets of H/ACA ribonucleoproteins. *Chromosoma* **114**, 1–14 (2005).
- Torchet, C. *et al.* The complete set of H/ACA snoRNAs that guide rRNA pseudouridylations in *Saccharomyces cerevisiae*. *RNA* **11**, 928–938 (2005).
- Lestrade, L. & Weber, M. J. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.* **34**, D158–D162 (2006).
- Ganot, P., Bortolin, M. L. & Kiss, T. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* **89**, 799–809 (1997).
- Ni, J., Tien, A. L. & Fournier, M. J. Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell* **89**, 565–573 (1997).
- Balakin, A. G., Smith, L. & Fournier, M. J. The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell* **86**, 823–834 (1996).
- Ganot, P., Caizergues-Ferrer, M. & Kiss, T. The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev.* **11**, 941–956 (1997).
- Rozhdetsvensky, T. S. *et al.* Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res.* **31**, 869–877 (2003).
- Morrissey, J. P. & Tollervey, D. Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis. *Mol. Cell. Biol.* **13**, 2469–2477 (1993).
- Atzorn, V., Fragapane, P. & Kiss, T. U17/snR30 is a ubiquitous snoRNA with two conserved sequence motifs essential for 18S rRNA production. *Mol. Cell. Biol.* **24**, 1769–1778 (2004).
- Mitchell, J. R., Cheng, J. & Collins, K. A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. *Mol. Cell. Biol.* **19**, 567–576 (1999).
- Girard, J. P. *et al.* GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast. *EMBO J.* **11**, 673–682 (1992).
- Bousquet-Antonelli, C., Henry, Y., G'elugne, J. P., Caizergues-Ferrer, M. & Kiss, T. A small nucleolar RNP protein is required for pseudouridylation of eukaryotic ribosomal RNAs. *EMBO J.* **16**, 4770–4776 (1997).
- Lafontaine, D. L., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M. & Tollervey, D. The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev.* **12**, 527–537 (1998).
- Watkins, N. J. *et al.* Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA* **4**, 1549–1568 (1998).
- Henras, A. *et al.* Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *EMBO J.* **17**, 7078–7090 (1998).
- Baker, D. L. *et al.* RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes Dev.* **19**, 1238–1248 (2005).
- Charpentier, B., Muller, S. & Branlant, C. Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res.* **33**, 3133–3144 (2005).
- Wang, C. & Meier, U. T. Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *EMBO J.* **23**, 1857–1867 (2004).
- Henras, A. K., Capeyrou, R., Henry, Y. & Caizergues-Ferrer, M. Cbf5p, the putative pseudouridine synthase of H/ACA-type snoRNPs, can form a complex with Gar1p and Nop10p in absence of Nhp2p and box H/ACA snoRNAs. *RNA* **10**, 1704–1712 (2004).
- Rashid, R. *et al.* Crystal structure of a Cbf5–Nop10–Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Mol. Cell* **21**, 249–260 (2006).
- Hamma, T., Reichow, S. L., Varani, G. & Ferre-D'Amare, A. R. The Cbf5–Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nature Struct. Mol. Biol.* **12**, 1101–1107 (2005).
- Manival, X. *et al.* Crystal structure determination and site-directed mutagenesis of the *Pyrococcus abyssi* aCBF5–aNOPI0 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity. *Nucleic Acids Res.* **34**, 826–839 (2006).
- Girard, J. P., Bagni, C., Caizergues-Ferrer, M., Amalric, F. & Lapeyre, B.

- Identification of a segment of the small nucleolar ribonucleoprotein-associated protein GAR1 that is sufficient for nucleolar accumulation. *J. Biol. Chem.* **269**, 18499–18506 (1994).
27. Ishitani, R. *et al.* Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme. *Cell* **113**, 383–394 (2003).
 28. Liang, X. H. *et al.* A genome-wide analysis of C/D and H/ACA-like small nucleolar RNAs in *Trypanosoma brucei* reveals a trypanosome-specific pattern of rRNA modification. *RNA* **11**, 619–645 (2005).
 29. Russell, A. G., Schnare, M. N. & Gray, M. W. Pseudouridine-guide RNAs and other Cbf5p-associated RNAs in *Euglena gracilis*. *RNA* **10**, 1034–1046 (2004).
 30. Heiss, N. S. *et al.* X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nature Genet.* **19**, 32–38 (1998).
 31. Marrone, A., Walne, A. & Dokal, I. Dyskeratosis congenita: telomerase, telomeres and anticipation. *Curr. Opin. Genet. Dev.* **15**, 249–257 (2005).
 32. Hoang, C. & Ferre-D'Amare, A. R. Cocystal structure of a tRNA^{Psi55} pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell* **107**, 929–939 (2001).
 33. Pan, H., Agarwalla, S., Moustakas, D. T., Finer-Moore, J. & Stroud, R. M. Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit. *Proc. Natl Acad. Sci. USA* **100**, 12648–12653 (2003).
 34. DeLano, W. L. *The PyMOL User's Manual* (Delano Scientific, San Carlos, California, 2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to D. J. Patel for his steady interest and encouragement in this project. We thank D. Patel and A. Serganov for critically reading the paper, and J. Duan, Y. Kang and W. Wang for assistance and discussion. We thank N. Shimizu and M. Kawamoto for help at SPring-8 beamline BL41XU. This research is funded by the Chinese Ministry of Science and Technology.

Author Contributions L.L. was responsible for biochemical and crystallization experiments; K.Y. determined the structure and wrote the paper. Both authors designed experiments and discussed results.

Author Information Coordinates and structure factors have been deposited in the Protein Data Bank under the accession code 2HVY. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.Y. (yekeqiong@nibs.ac.cn).

LETTERS

The type Ia supernova SNLS-03D3bb from a super-Chandrasekhar-mass white dwarf star

D. Andrew Howell¹, Mark Sullivan¹, Peter E. Nugent², Richard S. Ellis³, Alexander J. Conley¹, Damien Le Borgne⁴, Raymond G. Carlberg¹, Julien Guy⁵, David Balam⁶, Stephane Basa⁷, Dominique Fouchez⁸, Isobel M. Hook⁹, Eric Y. Hsiao⁶, James D. Neill⁶, Reynald Pain⁵, Kathryn M. Perrett¹ & Christopher J. Pritchet⁶

The accelerating expansion of the Universe, and the need for dark energy, were inferred from observations^{1,2} of type Ia supernovae. There is a consensus that type Ia supernovae are thermonuclear explosions that destroy carbon–oxygen white dwarf stars that have accreted matter from a companion star³, although the nature of this companion remains uncertain. These supernovae are thought to be reliable distance indicators because they have a standard amount of fuel and a uniform trigger: they are predicted to explode when the mass of the white dwarf nears the Chandrasekhar mass⁴ of 1.4 solar masses ($M_{\odot}$). Here we show that the high-redshift supernova SNLS-03D3bb has an exceptionally high luminosity and low kinetic energy that both imply a super-Chandrasekhar-mass progenitor. Super-Chandrasekhar-mass supernovae should occur preferentially in a young stellar population, so this may provide an explanation for the observed trend that overluminous type Ia supernovae occur only in ‘young’ environments^{5,6}. As this supernova does not obey the relations that allow type Ia supernovae to be calibrated as standard candles, and as no counterparts have been found at low redshift, future cosmology studies will have to consider possible contamination from such events.

The supernova SNLS-03D3bb (SN 2003fg) was discovered on 24 April 2003 UT as part of the Supernova Legacy Survey (SNLS). Its redshift is $z = 0.2440 \pm 0.0003$, determined from host galaxy [O II], [O III], H α , and H β emission lines. A finding chart and observational details can be found in the Supplementary Information. From the lightcurve (Fig. 1) we measure a peak magnitude in the rest frame V band, $V = 20.50 \pm 0.06$ mag. This corresponds to an absolute magnitude of $M_V = -19.94 \pm 0.06$ ($H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 0.3$, flat Universe). SNLS-03D3bb falls completely outside the M_V distribution of low- z type Ia supernovae (ref. 7), and is 0.87 mag (a factor of 2.2) brighter than the median. We note that neither changes in the Hubble constant nor Ω_M significantly affect this brightness difference. Asphericity may account for variations in luminosity in a type Ia supernova at the 25% level, but not by a factor of two^{8,9}. SNLS-03D3bb also does not follow the lightcurve width–luminosity relationship¹⁰ for type Ia supernovae that allows them to be calibrated as standard candles—it is too bright for its lightcurve width (‘stretch’, $s = 1.13$) by 0.61 ± 0.14 mag (4.4σ).

Type Ia supernovae are powered exclusively by the decay of ^{56}Ni and its decay product ^{56}Co (ref. 11), requiring $\sim 0.6 M_{\odot}$ of ^{56}Ni to reproduce a normal type Ia supernova^{12–15}. Because SNLS-03D3bb is 2.2 times overluminous, we infer that it has $\sim 1.3 M_{\odot}$ of ^{56}Ni . Such a

large ^{56}Ni mass is not possible if the progenitor is limited to the Chandrasekhar mass. Even models that burn the entire $1.4 M_{\odot}$ to nuclear statistical equilibrium via a pure detonation produce only $0.92 M_{\odot}$ of ^{56}Ni , with the remainder comprising other iron-peak elements¹⁶. At least 40% of the type Ia supernova must be elements other than ^{56}Ni to reproduce observed spectra^{14,17}; this implies a white-dwarf mass of $\sim 2.1 M_{\odot}$. Some authors find that rapid rotation may support such a massive white dwarf¹⁸. The merger of two massive white dwarfs could also produce a super-Chandrasekhar-mass product^{19,20}.

This simple estimation of the nickel mass is supported by a more

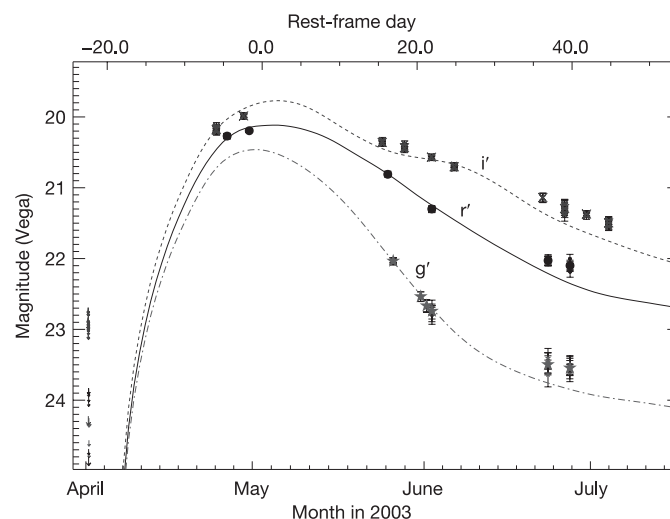


Figure 1 | The lightcurve of SNLS-03D3bb. We fitted k -corrected²⁷ template lightcurves to the observed photometry of SNLS-03D3bb, then transformed the peak magnitudes back to the Johnson–Cousins²⁸ BV magnitudes in the Vega system. We found peak magnitudes of $B = 20.35$ mag and $V = 20.50 \pm 0.06$ mag from a simultaneous fit to g' and r' data. The error (s.d.) consists of 0.04 statistical error and 0.04 k -correction error. A lightcurve template was fitted using the stretch method¹ (stretching the time axis of a template lightcurve by a stretch factor of $s = 1.13$). The epoch of maximum light relative to the rest-frame B band was determined from a simultaneous fit to all of the data. At maximum, we use only the V-band value to compare to other supernovae, because it is the best constrained. Data past +35 days were not used in the fit. The arrows are 3σ upper limits.

¹Department of Astronomy and Astrophysics, University of Toronto, 60 St George Street, Toronto, Ontario M5S 3H8, Canada. ²Lawrence Berkeley National Laboratory, Mail Stop 50-232, 1 Cyclotron Road, Berkeley, California 94720 USA. ³California Institute of Technology, East California Blvd, Pasadena, California 91125, USA. ⁴DAPNIA/Service d'Astrophysique, CEA/Saclay, 91191 Gif-sur-Yvette Cedex, France. ⁵LPNHE, CNRS-IN2P3 and University of Paris VI and VII, 75005 Paris, France. ⁶Department of Physics and Astronomy, University of Victoria, P O Box 3055, Victoria, British Columbia V8W 3P6, Canada. ⁷LAM CNRS, BP8, Traverse du Siphon, 13376 Marseille Cedex 12, France. ⁸CPPM, CNRS-IN2P3 and University Aix Marseille II, Case 907, 13288 Marseille Cedex 9, France. ⁹University of Oxford Astrophysics, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK.

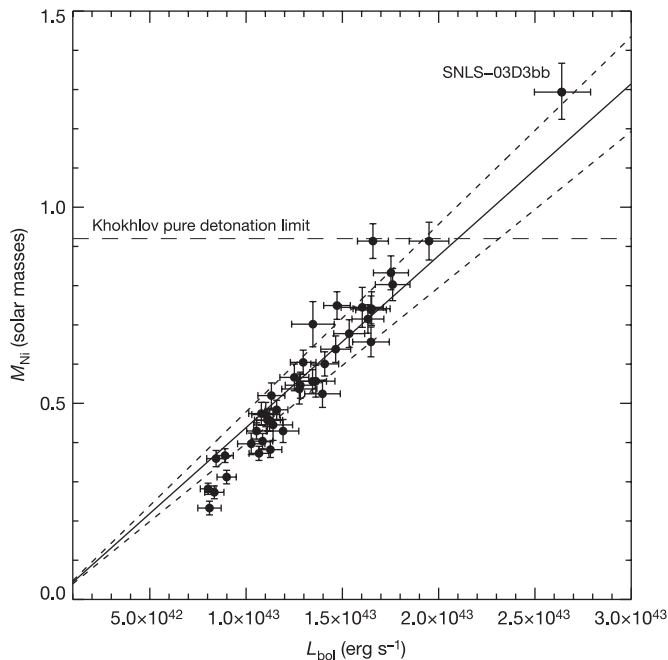


Figure 2 | Bolometric luminosity versus implied ^{56}Ni mass for SNLS-03D3bb and low-redshift type Ia supernovae. The low-redshift type Ia supernovae were fitted using the same techniques as those used for SNLS-03D3bb: the bolometric luminosity was determined using the peak magnitude in the V band from a simultaneous fit to B- and V-band data. For the low-redshift type Ia supernovae, we integrated the $s = 1$ supernova template²⁷ to obtain a bolometric correction of 0.06 mag and an uncertainty (s.d.) of 0.05 mag for the combined bolometric and k -correction error. The solid line represents a normal $s = 1$ type Ia supernova, with a rise time (t_R) of -19.5 days, while dotted lines show $s = 0.9$ ($t_R = 17.6$) and $s = 1.1$ ($t_R = 21.5$). Low-luminosity type Ia supernovae have lower stretches, and thus shorter rise times, resulting in less ^{56}Ni for a given luminosity, while high-stretch type Ia supernovae show opposite behaviour. The dashed line shows an upper limit for the expected ^{56}Ni mass in a Chandrasekhar-mass type Ia supernova, obtained by burning the entire white dwarf to iron-peak elements in a detonation¹⁶.

detailed calculation using the principle that the luminosity at maximum light is proportional to the instantaneous rate of radioactive decay^{21,22}. The implied Ni mass is^{23,24}: $M_{\text{Ni}} = \frac{L_{\text{bol}}}{\alpha \dot{S}(t_R)}$, where L_{bol} is the bolometric luminosity at maximum light (the luminosity integrated from the ultraviolet to the infrared), and α is the ratio of bolometric to radioactivity luminosities, near unity. $\dot{S}$ is the radioactivity luminosity per solar mass of ^{56}Ni from its decay to ^{56}Co and subsequent decay to ^{56}Fe : $\dot{S} = (6.31 \times 10^{43} e^{-t_R/8.8}) + (1.43 \times 10^{43} e^{-t_R/111}) \text{ erg s}^{-1} M_{\odot}^{-1}$, where t_R is the time in days for the supernova to rise from explosion to maximum light. Using $t_R = s \times 19.5$ days (ref. 25), for SNLS-03D3bb, we obtain $t_R = 22$ days (see Supplementary Information for the effect of a shorter rise). We use $\alpha = 1.2$ as a conservative value, although for high ^{56}Ni masses, α may be lower, because nickel above the photosphere will not contribute to the luminosity²³.

To convert our V magnitude into a bolometric equivalent, we use a synthetic spectrum calculated to match the observed ultraviolet + optical spectrum (Fig. 3), but extended into the infrared (13% of the bolometric luminosity is from the infrared extrapolation). The bolometric correction (m_{bc}) is 0.07 ± 0.03 mag, such that $M_{\text{bol}} = M_V + m_{\text{bc}} = -19.87 \pm 0.06$ mag. Using these numbers, we calculate $M_{\text{Ni}} = 1.29 \pm 0.07 M_{\odot}$ for SNLS-03D3bb, in agreement with the simple scaling argument used earlier. The quoted error is from the statistical, k -correction, and bolometric correction errors added in quadrature. SNLS-03D3bb has a significantly larger bolometric luminosity and implied ^{56}Ni mass compared to low-redshift supernovae (Fig. 2).

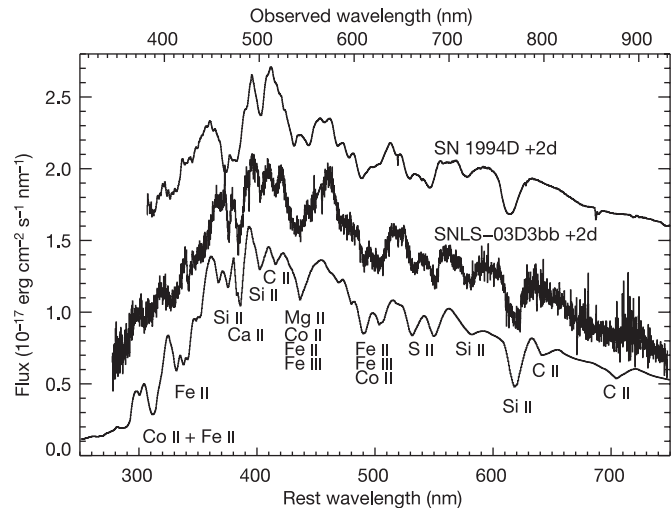


Figure 3 | Keck LRIS spectrum of SNLS-03D3bb at two days after maximum light compared to a spectrum of the normal type Ia supernova SN 1994D. Also plotted is a SYNOW fit to the data, with dominant ions labelled. SYNOW is a parameterized resonance-scattering code, allowing the user to adjust optical depths, temperatures, and velocities to aid in the identification of supernova lines²⁹. SYNOW parameters are listed in the Supplementary Information. SNLS-03D3bb shows the lines of intermediate-mass elements typically seen in a type Ia supernova at maximum light—Si II, S II, and Ca II,—but in SNLS-03D3bb the velocity of the lines is lower than usual. The line at 415 nm appears to be C II, but the other predicted carbon features cannot be clearly identified owing to the lower signal-to-noise ratio of the spectrum in the red. No other identification could be found for 415-nm feature.

SNLS-03D3bb also has an unusually low ejecta velocity, as shown in the Keck spectrum taken two days after maximum light (Fig. 3). With a Si II velocity of $8,000 \pm 500 \text{ km s}^{-1}$, it falls well outside the range of velocities seen for this feature at maximum light (Fig. 4). This is hard to understand in the Chandrasekhar-mass model, which predicts higher velocities for more luminous type Ia supernovae, in contrast to the unusually low velocities in SNLS-03D3bb.

The kinetic energy (E_K) of a type Ia supernova arises from the difference between the nuclear energy (E_N) obtained from the synthesis of elements via fusion in the explosion and the binding energy (E_b) of the white dwarf¹³. Thus the kinetic energy velocity is: $v_{\text{ke}} = \sqrt{\frac{2(E_N - E_b)}{M_{\text{wd}}}}$, where M_{wd} is the mass of the white dwarf. The binding energy of a $1.4 M_{\odot}$ carbon–oxygen white dwarf is $0.5 \times 10^{51} \text{ ergs}$ (ref. 14). For a $2 M_{\odot}$ white dwarf and a central density of $4 \times 10^9 \text{ g cm}^{-3}$, the binding energy is $1.3 \times 10^{51} \text{ ergs}$ (ref. 18).

There are only three classes of elements in a type Ia supernova (iron-peak elements, intermediate mass elements, and unburned carbon and oxygen), so a simple model can be developed for the nuclear-energy generation, E_N . Burning a mixture of equal parts carbon and oxygen to the iron peak produces $E_{\text{Fe}} = 1.55 \times 10^{51} \text{ erg } M_{\odot}^{-1}$, while the synthesis of ^{28}Si produces 76% as much energy¹³. Thus: $E_N = E_{\text{Fe}} M_{\text{wd}} (f_{\text{Fe}} + 0.76 f_{\text{IME}})$, where M_{wd} is in solar masses, and f_{Fe} and f_{IME} are the fractional compositions of iron peak and intermediate-mass elements. If f_{Fe} and f_{IME} do not sum to one, the remainder is the fraction of unburned carbon and oxygen (f_C), which does not contribute to the nuclear energy. The ^{56}Ni makes up approximately 70% of iron-peak elements^{14,16}, so we adopt $M_{\text{Ni}} = 0.7 M_{\text{wd}} f_{\text{Fe}}$, where M_{Ni} is the mass of ^{56}Ni .

In the Chandrasekhar-mass model, more-luminous supernovae, with more ^{56}Ni , have a higher v_{ke} (Fig. 4). Increasing the fraction of unburned carbon and oxygen, f_C , can lower the kinetic energy, perhaps accounting for some of the dispersion in type Ia supernova velocities. However, this also lowers the available ^{56}Ni , so it cannot account for the low velocity seen in SNLS-03D3bb.

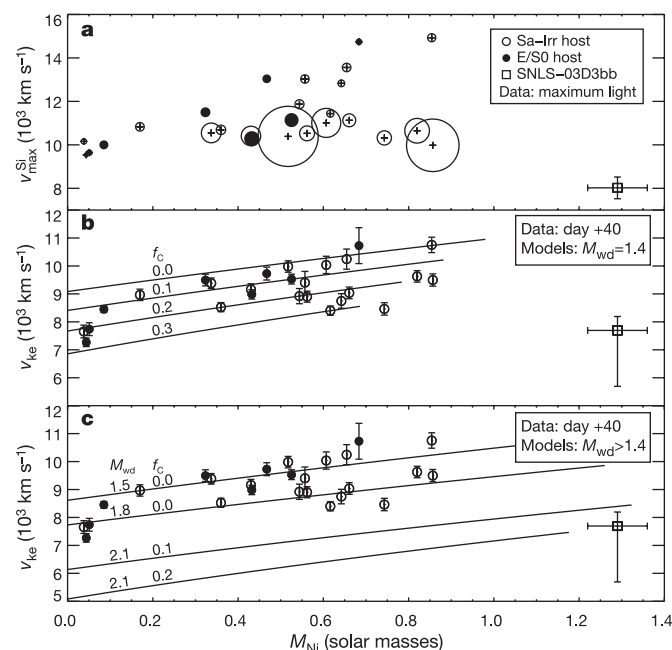


Figure 4 | Inferred Ni mass versus Si II 615 nm velocity. **a**, The data at maximum light²⁶. Ni masses are converted as described in the text using $M_{\text{bol}} = M_{\text{B}} + 0.2$. Filled circles are from early-type galaxies (E or S0), while open circles are from late-type galaxies (Sa–Irr). Circle size is proportional to $v_{\text{Si}}/\dot{v}_{\text{Si}}$, where $\dot{v}_{\text{Si}}$ is the rate of change of the velocity of the Si II feature. There is no $\dot{v}_{\text{Si}}$ measurement for SNLS-03D3bb. **b**, Kinetic-energy velocity of type Ia supernovae versus Ni mass for $1.4M_{\odot}$ models with different fractions of unburned carbon (f_{C}). This unburned fraction should not be much higher than $\sim 20\%$ because carbon is rarely seen in type Ia supernova spectra³⁰. Overplotted symbols are v_{Si} for low-redshift type Ia supernovae²⁶ extrapolated to 40 days after maximum (correcting for stretch). For SNLS-03D3bb we use $\dot{v}_{\text{Si}}$ from its closest neighbour. The error bar reflects the range if an average value of $\dot{v}_{\text{Si}}$ is used. SNLS-03D3bb is not consistent with the $1.4M_{\odot}$ model. **c**, As above, but showing that $M_{\text{wd}} \approx 2M_{\odot}$ models can explain SNLS-03D3bb. Less extreme super-Chandrasekhar-mass models are consistent with the low-redshift data. The three low ^{56}Ni supernovae are not necessarily of super-Chandrasekhar mass—their large values of $\dot{v}_{\text{Si}}$ make projections to 40 days uncertain.

The kinetic energy gives the velocity of the supernova averaged over the entire mass, approximately equivalent to the velocity a few weeks after maximum light¹³. The most appropriate observational signature of this velocity is unclear, because type Ia supernova line velocities change with time, and different ions can have different relative velocities. We find good agreement between the Si II velocity at 40 days after maximum light²⁶ and the theoretical kinetic-energy velocity, but we emphasize that this is an imperfect comparison.

A super-Chandrasekhar mass reproduces the low velocities seen in SNLS-03D3bb (Fig. 4). Because Chandrasekhar models with more Ni produce higher velocities, the low velocities of SNLS-03D3bb imply an increased progenitor binding energy and thus a larger total mass. As a caveat, we note that this simple calculation is only intended to illustrate general trends. Future theoretical studies will have to assess such complications using different ions, different white-dwarf density structures, and a wider range of binding energies.

Super-Chandrasekhar-mass type Ia supernovae should be more likely to occur in a young stellar population, where the most massive stars exist^{19,20}. The low-mass, star-forming host of SNLS-03D3bb is consistent with this scenario (see Supplementary Information). Thus, the apparent existence of super-Chandrasekhar-mass type Ia supernovae may explain why the most luminous type Ia supernovae only

occur in young stellar environments^{5,6}. The standard Chandrasekhar-mass model offers no explanation for this behaviour, because the total amount of fuel and triggering mechanism should be independent of the mass of the progenitor stars.

Supernovae such as SNLS-03D3bb will have to be screened out in cosmological studies. Younger stellar environments produce more-luminous supernovae, so as the mean stellar age decreases with redshift the mean properties of type Ia supernovae will change⁵. This can be calibrated if all supernovae obey the same stretch–luminosity relationship, but SNLS-03D3bb does not. Its peculiarity was so obvious that it was excluded from the SNLS cosmological result⁷, but less extreme objects could lurk in supernova samples. Future cosmology studies will need to carefully scrutinise type Ia supernovae from young populations to see whether they obey the same lightcurve width–luminosity relationship as other type Ia supernovae.

Received 7 April; accepted 18 July 2006.

1. Perlmutter, S. *et al.* Measurements of Omega and Lambda from 42 high-redshift supernovae. *Astrophys. J.* **517**, 565–586 (1999).
2. Riess, A. G. *et al.* Observational evidence from supernovae for an accelerating Universe and a cosmological constant. *Astron. J.* **116**, 1009–1038 (1998).
3. Hoyle, F. & Fowler, W. A. Nucleosynthesis in supernovae. *Astrophys. J.* **132**, 565–590 (1960).
4. Chandrasekhar, S. The maximum mass of ideal white dwarfs. *Astrophys. J.* **74**, 81–82 (1931).
5. Sullivan, M. *et al.* Rates and properties of type Ia supernovae as a function of mass and star-formation in their host galaxies. *Astrophys. J.* (in the press); preprint at (<http://www.arXiv.org/astro-ph/0605455>) (2006).
6. Hamuy, M. *et al.* The absolute luminosities of the Calan/Tololo type Ia supernovae. *Astron. J.* **112**, 2391–2397 (1996).
7. Astier, P. *et al.* The Supernova Legacy Survey: measurement of Ω_{M} , Ω_{Λ} and w from the first year data set. *Astron. Astrophys.* **447**, 31–48 (2006).
8. Howell, D. A., Höflich, P., Wang, L. & Wheeler, J. C. Evidence for asphericity in a subluminescent type Ia supernova: spectropolarimetry of SN 1999by. *Astrophys. J.* **556**, 302–321 (2001).
9. Kasen, D., Nugent, P., Thomas, R. C. & Wang, L. Could there be a hole in type Ia supernovae? *Astrophys. J.* **610**, 876–887 (2004).
10. Phillips, M. M. The absolute magnitudes of type Ia supernovae. *Astrophys. J. Lett.* **413**, L105–L108 (1993).
11. Colgate, S. A. & McKee, C. Early supernova luminosity. *Astrophys. J.* **157**, 623–644 (1969).
12. Woosley, S. E. & Weaver, T. A. The physics of supernova explosions. *Annu. Rev. Astron. Astrophys.* **24**, 205–253 (1986).
13. Branch, D. The Hubble constant from nickel radioactivity in type Ia supernovae. *Astrophys. J.* **392**, 35–40 (1992).
14. Nomoto, K., Thielemann, F.-K. & Yokoi, K. Accreting white dwarf models of Type I supernovae. III—Carbon deflagration supernovae. *Astrophys. J.* **286**, 644–658 (1984).
15. Höflich, P. Analysis of the type Ia supernova SN 1994D. *Astrophys. J.* **443**, 89–108 (1995).
16. Khokhlov, A., Mueller, E. & Höflich, P. Light curves of type Ia supernova models with different explosion mechanisms. *Astron. Astrophys.* **270**, 223–248 (1993).
17. Fisher, A., Branch, D., Hatano, K. & Baron, E. On the spectrum and nature of the peculiar type Ia supernova 1991T. *Mon. Not. R. Astron. Soc.* **304**, 67–74 (1999).
18. Yoon, S.-C. & Langer, N. On the evolution of rapidly rotating massive white dwarfs towards supernovae or collapses. *Astron. Astrophys.* **435**, 967–985 (2005).
19. Tutukov, A. V. & Yungelson, L. R. Merging of binary white dwarfs neutron stars and black-holes under the influence of gravitational wave radiation. *Mon. Not. R. Astron. Soc.* **268**, 871–879 (1994).
20. Howell, D. A. The progenitors of subluminescent type Ia supernovae. *Astrophys. J. Lett.* **554**, L193–L196 (2001).
21. Arnett, W. D. Type I supernovae. I—Analytic solutions for the early part of the light curve. *Astrophys. J.* **253**, 785–797 (1982).
22. Arnett, W. D., Branch, D. & Wheeler, J. C. Hubble's constant and exploding carbon-oxygen white dwarf models for type I supernovae. *Nature* **314**, 337–338 (1985).
23. Nugent, P. *et al.* Low Hubble constant from the physics of type Ia supernovae. *Phys. Rev. Lett.* **75**, 394–397 (1995).
24. Nugent, P. E. Non-local thermodynamic equilibrium spectrum synthesis of type Ia supernovae. PhD thesis, Univ. Oklahoma (1997).
25. Conley, A. *et al.* The rise time of type Ia supernovae from the Supernova Legacy Survey. *Astron. J.* (in the press); preprint at (<http://www.arXiv.org/astro-ph/0607363>) (2006).

26. Benetti, S. *et al.* The diversity of type Ia supernovae: evidence for systematics? *Astrophys. J.* **623**, 1011–1016 (2005).
27. Nugent, P., Kim, A. & Perlmutter, S. K-corrections and extinction corrections for type Ia supernovae. *Publ. Astron. Soc. Pacif.* **114**, 803–819 (2002).
28. Bessell, M. S. UBVRI passbands. *Publ. Astron. Soc. Pacif.* **102**, 1181–1199 (1990).
29. Fisher, A., Branch, D., Nugent, P. & Baron, E. Evidence for a high-velocity carbon-rich layer in the type Ia SN 1990N. *Astrophys. J. Lett.* **481**, L89–L92 (1997).
30. Marion, G. H. *et al.* Low carbon abundance in type Ia supernovae. *Astrophys. J.* **645**, 1392–1401 (2006).

Supplementary Information is linked to the online version of this paper at www.nature.com/nature.

Acknowledgements SNLS relies on observations with MegaCam, a joint project of CFHT and CEA/DAPNIA, at the Canada–France–Hawaii Telescope (CFHT). We used data products from the Canadian Astronomy Data Centre as part of the CFHT Legacy Survey. Some data were obtained at the W. M. Keck Observatory. We acknowledge support from NSERC, NERSC, CIAR, CNRS/IN2P3, CNRS/INSU, CEA and the DOE.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.A.H. (howell@astro.utoronto.ca).

LETTERS

Spontaneous symmetry breaking in a quenched ferromagnetic spinor Bose–Einstein condensate

L. E. Sadler¹, J. M. Higbie¹, S. R. Leslie¹, M. Vengalattore¹ & D. M. Stamper-Kurn¹

A central goal in condensed matter and modern atomic physics is the exploration of quantum phases of matter—in particular, how the universal characteristics of zero-temperature quantum phase transitions differ from those established for thermal phase transitions at non-zero temperature. Compared to conventional condensed matter systems, atomic gases provide a unique opportunity to explore quantum dynamics far from equilibrium. For example, gaseous spinor Bose–Einstein condensates^{1–3} (whose atoms have non-zero internal angular momentum) are quantum fluids that simultaneously realize superfluidity and magnetism, both of which are associated with symmetry breaking. Here we explore spontaneous symmetry breaking in ⁸⁷Rb spinor condensates, rapidly quenched across a quantum phase transition to a ferromagnetic state. We observe the formation of spin textures, ferromagnetic domains and domain walls, and demonstrate phase-sensitive *in situ* detection of spin vortices. The latter are topological defects resulting from the symmetry breaking, containing non-zero spin current but no net mass current⁴.

Spinor atomic gases^{1–3} are those comprised of atoms with non-zero internal angular momentum—the sum of electronic and nuclear angular momenta, denoted by quantum number F —and in which all orientations of the atomic spin may be realized. A spinor gas Bose–Einstein condensate (BEC) is described by a vector order parameter and therefore exhibits spontaneous magnetic ordering. Nevertheless, freedom remains for the type of ordering that can occur. For ⁸⁷Rb $F = 1$ spinor gases^{5,6}, the spin-dependent energy per particle in the condensate is the sum of two terms, $c_2 n \langle \hat{\mathbf{F}} \rangle^2 + q \langle \hat{F}_z^2 \rangle$, where $\hat{\mathbf{F}}$ denotes the dimensionless spin vector operator. The first term describes spin-dependent interatomic interactions, with n being the number density and $c_2 = (4\pi\hbar^2/3m)(a_2 - a_0)$ depending on the atomic mass m and the s -wave scattering lengths a_f for collisions between pairs of particles with total spin f (refs 2 and 3). Given $c_2 < 0$ for our system^{5–7}, the interaction term alone favours a ferromagnetic phase with broken rotational symmetry. The second term describes a quadratic Zeeman shift in our experiment, with $q = (\hbar \times 70 \text{ Hz G}^{-2}) B^2$ at a magnetic field of magnitude B (the linear Zeeman shift may be neglected owing to spin conservation). This term favours instead a phase with no net magnetization, that is, a condensate in the $|m_z = 0\rangle$ magnetic sublevel, with unbroken $O(2)$ rotational symmetry in the transverse spin plane. These two phases have distinct symmetries, and are therefore divided by a quantum phase transition at $q = 2|c_2|n$.

Here we describe the observation of spontaneous symmetry breaking in a ⁸⁷Rb spinor BEC that is rapidly quenched across this quantum phase transition. Nearly pure spinor BECs were prepared in the unmagnetized $|m_z = 0\rangle$ phase at a high quadratic Zeeman shift ($q \gg 2|c_2|n$). By rapidly reducing the magnitude of the applied magnetic field, we quenched the gas to conditions that favour the ferromagnetic phase ($q \ll 2|c_2|n$). At variable times T_{hold} after the

quench, high-resolution maps of the magnetization vector density were obtained using magnetization-sensitive phase-contrast imaging⁸. After the quench, transverse ferromagnetic domains of variable size formed spontaneously throughout the condensate, divided by narrow unmagnetized domain walls. Concurrent with the formation of these domains, we also observed topological defects that we characterize as singly charged spin vortices with circulating spin currents and unmagnetized filled cores.

Spinor BECs in the $|F = 1, m_z = 0\rangle$ hyperfine state were confined in an optical dipole trap characterized by oscillation frequencies $(\omega_x, \omega_y, \omega_z) = 2\pi(56, 350, 4.3) \text{ s}^{-1}$. The condensates, typically containing $2.1(1) \times 10^6$ atoms, were formed at a magnetic field of 2 G and characterized by a peak density $n_0 = 2.8 \times 10^{14} \text{ cm}^{-3}$ and Thomas–Fermi radii $(r_x, r_y, r_z) = (12.8, 2.0, 167) \mu\text{m}$ (see Methods). Variations in the internal-state wavefunction were constrained in these anisotropic condensates to just two spatial dimensions ($\hat{x}$ and $\hat{z}$) because the spin healing length, $\xi_s = \sqrt{\hbar^2/2m|c_2|n_0} = 2.4 \mu\text{m}$, was larger than the cloud size r_y in the $\hat{y}$ direction. Thus, imaging the condensate in the $\hat{x}$ – $\hat{z}$ plane produced complete maps of the magnetization density.

After the condensate was formed, the magnetic field was oriented in the $\hat{z}$ direction, ramped linearly over 5 ms to a magnitude of 50 mG, and held at this setting for a variable time T_{hold} before we imaged the gas. At this field, the quadratic Zeeman energy $q = \hbar \times 0.2 \text{ Hz}$ is negligible compared to twice the spin-dependent interaction energy of $2|c_2|\langle n \rangle_y = \hbar \times 16 \text{ Hz}$, where $\langle n \rangle_y$ is the density averaged in the $\hat{y}$ direction.

The condensate magnetization was measured *in situ* using phase-contrast imaging, which yields a magnetization-sensitive signal given approximately as ζF_y , where ζ is proportional to the gas column density and $F_y = \langle \hat{F}_y \rangle$ is one component of the (dimensionless) magnetization of the gas (see Methods). We determined all three components of the vector magnetization density with repeated imaging of the same atomic sample. Transverse magnetization was detected by imaging its Larmor precession about a $\hat{z}$ -oriented guide field⁸. The complex transverse magnetization $F_t = F_x + iF_y$ was then determined as $A(\boldsymbol{\rho})\exp(i\phi(\boldsymbol{\rho})) = i\zeta(\boldsymbol{\rho})F_t(\boldsymbol{\rho})$ from the amplitude $A(\boldsymbol{\rho})$ and phase of $\phi(\boldsymbol{\rho})$ of Larmor precession at each pixel position $\boldsymbol{\rho}$. Longitudinal magnetization was then measured from images in which the magnetic field was adiabatically reoriented in the $\pm\hat{y}$ directions.

As shown in Fig. 1, at short times after the quench ($T_{\text{hold}} < 50 \text{ ms}$), images probing the transverse magnetization show no significant variation across the cloud or between frames (similarly for the longitudinal magnetization images), indicating the presence of BECs remaining in the unmagnetized phase. Any magnetization during this stage was either too low in magnitude or varied over too short a length scale to be discerned by our imaging. At later times, a non-zero transverse magnetization signal spontaneously developed,

¹Department of Physics, University of California, Berkeley, California 94720, USA.

yielding a Larmor precession signal that varied both in amplitude and in phase across the condensate. This observation indicates heterogeneous, spontaneous symmetry breaking in the gas, specifically the breaking of $O(2)$ rotational symmetry in the transverse plane in a direction given by the phase of Larmor precession.

In Fig. 2, the derived transverse magnetization for samples at variable times T_{hold} are presented, rendered in colour to portray both the magnetization orientation (as hue) and amplitude (as brightness). These spatial maps show ferromagnetic domains of variable size and orientation arising spontaneously after the quench. The landscape of domains includes small regions of homogeneous magnetization, with unmagnetized domain walls separating regions of nearly opposite orientation, and also larger ferromagnetic spin textures free of domain walls in which the magnetization orientation varies smoothly over tens of micrometres.

Ferromagnetic domains from a quenched unmagnetized ($|m_z = 0\rangle$) gas arise through a dynamical instability in a spinor BEC^{9–12}. This instability, which is a consequence of coherent collisional mixing between magnetic sublevels^{5,6,13,14}, may also be regarded as the phase separation of a two-component condensate^{15,16}. For this, we recall that the $|m_z = 0\rangle$ state represents a coherent superposition of the $|m_\phi = \pm 1\rangle$ eigenstates of any transverse spin operator, $\hat{F}_\phi = \hat{F}_x \cos \phi + \hat{F}_y \sin \phi$, and that, because $c_2 < 0$, the ± 1 eigenstates of any spin component are immiscible¹⁷. Ferromagnetism thus arises by the spinodal decomposition of a binary gaseous mixture into neighbouring regions of oppositely oriented transverse magnetization. This phase separation is dominated by a fast-growing instability with an exponential timescale of $\tau_{\text{fm}} = \hbar / \sqrt{2|c_2|n}$. The wavevector of the dominant instability $k_{\text{fm}} = \sqrt{2m|c_2|n}/\hbar$ defines the typical size $l = \pi k_{\text{fm}}^{-1}$ of single-component domains in the phase-separated fluid, and also the width $b \approx k_{\text{fm}}^{-1}$ of domain walls in which the two components still overlap^{18,19}.

We emphasize that domain formation is neither the cause nor the mechanism for spontaneous symmetry breaking. In the context of mean-field theory, the above mentioned dynamical instability describes the amplification of an initial seed of non-zero transverse

magnetization of a particular spin orientation (denoted above as ϕ) once the rotational symmetry of the initial unmagnetized gas is already broken. The source of such a symmetry-broken seed is either thermal or quantum magnetization fluctuations in the gas before the quench. In this work, the forced depletion of atoms not in the $|m_z = 0\rangle$ state and the energy gap for magnetization fluctuations in the unmagnetized (high- q) phase suggest that ferromagnetism formed purely by the amplification of quantum fluctuations, that is, shot noise, a suggestion that warrants experimental justification.

To compare our data to the model of dynamical instabilities described above, we considered the density-weighted transverse magnetization correlation function:

$$G_t(\delta\mathbf{p}) = \text{Re} \left[\frac{\sum_{\mathbf{p}} (\zeta(\mathbf{p}) F_t(\mathbf{p}))^* (\zeta(\mathbf{p} + \delta\mathbf{p}) F_t(\mathbf{p} + \delta\mathbf{p}))}{\sum_{\mathbf{p}} \zeta(\mathbf{p}) \zeta(\mathbf{p} + \delta\mathbf{p})} \right]$$

At zero range, $G_t(0)$ measures the degree to which the condensate has evolved toward the ferromagnetic state. As shown in Fig. 3a, $G_t(0)$ rises after the quench from a near-zero value characteristic of the unmagnetized state to a nearly constant value of $G_t(0) \approx 0.5$. This saturation value measures the area occupied by domain walls. Fitting $G_t(0)$ at early times ($T_{\text{hold}} < 90$ ms) to a rising exponential yields a time constant of $\tau = 15(4)$ ms, in agreement with the predicted $\tau_{\text{fm}} = \hbar / \sqrt{2|c_2|n} = 13.7(3)$ ms. Over the same period of evolution, no significant longitudinal magnetization was observed, confirming the presence of purely transverse ferromagnetic domains.

Spatial correlations in the transverse magnetization (Fig. 3b) are

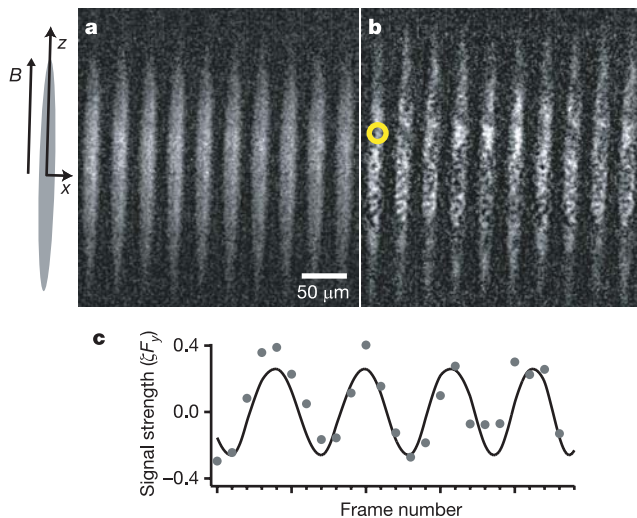


Figure 1 | Direct imaging of inhomogeneous spontaneous magnetization of a spinor BEC. Transverse imaging sequences (first 10 of 24 frames) are shown for a single condensate probed at $T_{\text{hold}} = 36$ ms (a) and for a different condensate at $T_{\text{hold}} = 216$ ms (b). Shortly after the quench, the system remains unmagnetized, showing neither short-range spatial nor temporal variation (that is, between frames). In contrast, condensates at longer times are spatially inhomogeneous and display spontaneous Larmor precession. The precession signal from one pixel position (at centre of yellow circle in b) is presented in c with data (circles) and fitted precession (line) both shown. Orientations of axes and magnetic field are shown on the left.

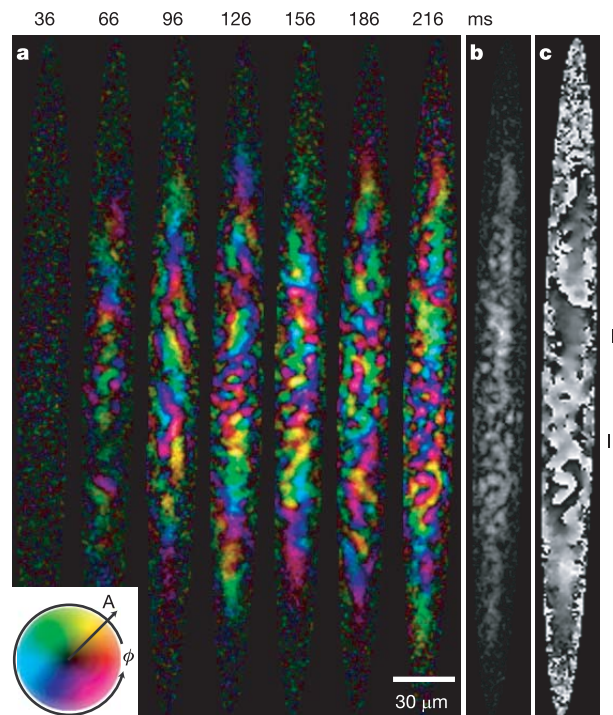


Figure 2 | In situ images of ferromagnetic domains and domain walls. a, The transverse magnetization density, measured for condensates at variable times T_{hold} (given on top), is shown with the magnetization orientation ($\phi = \arg(F_t)$) indicated by hue, and magnitude ($A = \zeta|F_t|$) by brightness. The maximum brightness, shown by the colour wheel on the left, corresponds to full magnetization of the condensate centre. For the data at $T_{\text{hold}} = 216$ ms, the magnetization density (b) and orientation (c) are shown separately. The spin texture at position I is characterized by an extended region (along $\hat{z}$) of non-zero magnetization magnitude, and a slowly varying magnetization orientation. Smaller magnetic domains at position II are divided by domain walls with zero magnetization. The greyscale in c covers the range 0 to 2π . Regions outside the condensate are coloured in black.

typified by a central region of positive correlations (near $\delta\mathbf{p} = 0$) and then several equally sized regions of alternating negative and positive correlations displaced from one another in the narrow $\hat{x}$ dimension of the condensate. We estimate a typical size for single-component domains as $\approx 10\ \mu\text{m}$, twice the displacement at which the transverse spin-spin correlation changes sign, in good agreement with the predicted $\pi k_{\text{fm}}^{-1} = 8.3(2)\ \mu\text{m}$. The presence of negative correlation regions supports the model of spin-conserving phase separation discussed above. The preferential phase separation in the narrow $\hat{x}$ direction rather than along $\hat{z}$, as indicated by $G_t(\delta\mathbf{p})$, may arise because the momentum distribution of the unmagnetized condensate in that direction is broader owing to the finite condensate size. Thus, upon quenching the system, a greater population of atoms is available to seed the faster-growing, shorter-wavelength instabilities with wavevector in the $\hat{x}$ direction. The presence of several alternations of positive and negative correlations further suggests that the phase separation occurs through a small number of discrete, unstable magnetization modes.

The spontaneous symmetry breaking observed in this work is one of many examples of symmetry breaking that occur in nature. For example, symmetry breaking is presumed to have occurred at thermal phase transitions in the early Universe, giving rise to the specific elementary particles and interactions now observed. An important aspect of rapid spontaneous symmetry breaking, whether in the laboratory or of a cosmological nature, is the creation of topological defects^{20–22}. The types of topological defects that may be formed depend on the group structure of the ground-state manifold reached at the transition.

Analogues of cosmological topological defect formation have been studied using liquid crystals²³ and superfluid helium^{24–26}. In

comparison with these previous experiments, the present investigation focuses on a simpler physical system, in which the time for thermal equilibration and domain formation is much longer than that needed to bring the system across the symmetry-breaking transition. Furthermore, the present work focuses on a quantum rather than a thermal phase transition²².

In our two-dimensional system, spin vortices are topological point defects about which the orientation of magnetization has a $2\pi l$ winding with l being a non-zero integer. Spin-vortex defects were observed with high confidence in about one-third of all images containing significant ferromagnetism ($T_{\text{hold}} > 90\ \text{ms}$), with some images indicating as many as four vortices (see Methods). Data revealing one such spin vortex are shown in Fig. 4. For this vortex, the central region of near-zero Larmor precession amplitude has a diameter of about $3\ \mu\text{m}$, comparable to the spin healing length $\xi_s = 2.4\ \mu\text{m}$ defined earlier. All observed vortices were singly quantized, with no apparent preferred direction of circulation.

Many types of vortices that can occur in a gas with a multi-component order parameter can be distinguished by the composition of their cores²⁷. On the basis of our measurements of the transverse and longitudinal magnetization at the vortex core, the spin vortices seen in our experiment appear to have unmagnetized filled cores. This observation supports their characterization as ‘polar-core’ spin vortices (denoted as $(\pm 1, 0, \mp 1)$ vortices in ref. 27), for which the superfluid order parameter is a superposition of atoms in the $|m_z = 1\rangle$ state rotating with one quantum of circulation, atoms in the $|m_z = -1\rangle$ state rotating with one quantum of circulation in the opposite sense, and non-rotating atoms in the unmagnetized $|m_z = 0\rangle$ state, which also fill the vortex core. Such a vortex is thus characterized by zero net mass circulation and a spin current with one quantum of circulation. The origin for such spin currents is presumably the spinodal decomposition by which ferromagnetism emerges from the unmagnetized cloud. The generation of such vortices under conditions of our experiment was predicted in ref. 4. The alternate identification of these vortices as merons²⁸ is ruled out by the absence of longitudinal magnetization at the vortex core.

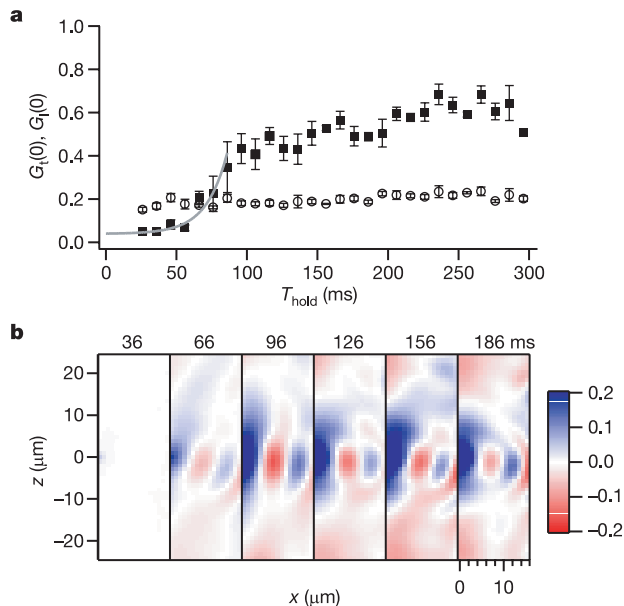


Figure 3 | Temporal and spatial evolution of ferromagnetism in a quenched spinor BEC. **a**, The mean local squared transverse ($G_t(0)$, squares) and longitudinal ($G_l(0)$, circles) magnetization, averaged over several experimental repetitions (error bars indicate shot-to-shot root-mean-square fluctuations), is shown. $G_t(0)$ rises exponentially with time constant $\tau = 15(4)\ \text{ms}$ as determined by a fit to data for $T_{\text{hold}} < 90\ \text{ms}$ (grey line), and saturates at $G_t(0) \approx 0.5$ at later times. No significant longitudinal magnetization is observed. The minimum values for $G_t(0)$ and $G_l(0)$, evident in the data for earliest T_{hold} , reflect residual noise due to processing either 24 or 2 frames, respectively, to obtain the magnetization density. **b**, Spatial correlation in the transverse magnetization ($G_t(x, z)$), at variable T_{hold} shows alternating regions of positive and negative correlations in the narrow $\hat{x}$ direction. At each T_{hold} , $G_t(x, z)$ was determined for ten experimental repetitions and then averaged.

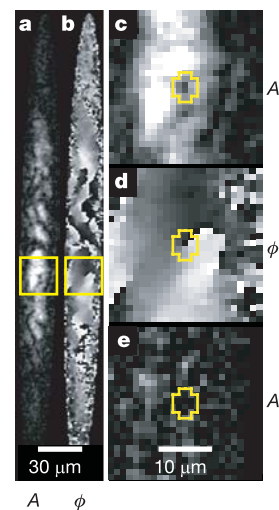


Figure 4 | In situ detection of a polar-core spin vortex. Spatial maps of the transverse magnetization magnitude ($A = |F_t|$) (**a**) and orientation ($\phi = \arg(F_t)$) (**b**) are shown for a sample imaged at $T_{\text{hold}} = 156\ \text{ms}$. Data from a portion of the condensate, indicated by boxes, are magnified, showing the transverse magnetization magnitude (**c**) and orientation (**d**) and also the magnitude of the longitudinal magnetization ($A_l = |F_z|$) (**e**). The phase along a closed path (indicated in yellow) surrounding a region of near-zero transverse magnetization shows a net winding of 2π , revealing the presence of a spin-vortex defect. The core (within the closed path) shows no significant longitudinal magnetization, allowing the identification of the defect as a polar-core spin vortex. The grayscale for images **a**, **c** and **e** is the same.

In contrast with their topological nature in some other magnetic systems, domain walls in a $F = 1$ ferromagnet are not topologically stable; rather, they may decay by the formation of spin vortex–antivortex pairs. However, we observe domain walls that form at the onset of visible ferromagnetism in the gas and persist for all times thereafter. In future experiments, our non-destructive imaging method may be adapted to allow the behaviour of a single quenched condensate to be recorded, allowing a closer examination of the dynamical evolution of domain walls. Improved magnetization imaging may allow studies of hourspin vortices are produced^{20,21} and of their subsequent evolution. Further time-resolved experiments in which one varies the rate at which the system is swept into the ferromagnetic state may also uncover universal temporal dynamics that typify this and other quantum phase transitions.

METHODS

Experimental sequence. Optically trapped BECs were obtained by loading around 10^8 atoms in the $|F = 1, m_z = -1\rangle$ state with temperature of $2.5\ \mu\text{K}$ into an optical dipole trap. The optical trap was formed by a single focused laser beam with a wavelength of $825\ \text{nm}$, which was linearly polarized to ensure that all components experience the same trap potential. A small shift quadratic in m_F from the optical beam was negligible.

Using radio-frequency rapid adiabatic passage followed by application of a transient magnetic field gradient of $4\ \text{G cm}^{-1}$, all atoms in the optical trap were placed in the $|m_z = 0\rangle$ magnetic sublevel. While the magnetic field was held at a magnitude of $2\ \text{G}$, the trap depth was decreased over $400\ \text{ms}$ to $k_B \times 350\ \text{nK}$ and the temperature of the gas to $\sim 40\ \text{nK}$, well below the Bose–Einstein condensation temperature. This yielded nearly pure condensates of $2.1(1) \times 10^6$ atoms.

At the low-field setting following the quench, magnetic field gradients along the $\hat{x}$ and $\hat{z}$ directions were nulled to less than $0.2\ \text{mG cm}^{-1}$. A Stern–Gerlach analysis of populations in each of the magnetic sublevels was applied to establish that the field ramp was sufficiently slow so as not to alter the spin state of the condensate. Such an analysis bounded the populations in each of the $|m_z = \pm 1\rangle$ spin states to be less than 0.3% of the total population, both before and right after the field ramp. At later times T_{hold} after the quench, such measurements showed significant mixing of populations coincident with the onset of spontaneous Larmor precession^{5,6}.

Magnetization imaging. The probe light for magnetization-sensitive phase-contrast imaging was circularly polarized and detuned by $\delta = -200\ \text{MHz}$ from the $5S_{1/2}(F = 1) \rightarrow 5P_{1/2}(F' = 2)$ transition. At this setting, the phase-contrast signal is given approximately as $\zeta(\frac{6}{5} + \langle \hat{F}_x \rangle + \frac{1}{5}\langle \hat{F}_y^2 \rangle)$ where $\zeta = (5/48)\tilde{n}\sigma\gamma/\delta$, $\tilde{n}$ is the column density of the gas, $\sigma = 3\lambda^2/2\pi$ is the resonant cross-section, $\lambda = 795\ \text{nm}$ is the wavelength of the probe light, and γ is the natural linewidth. We ignore the small $\langle \hat{F}_y^2 \rangle$ signal in this work. Probe pulses were $1\ \mu\text{s}$ in duration, much shorter than the $\sim 35\ \text{kHz}$ Larmor precession frequency at a $50\ \text{mG}$ field. The imaging system was diffraction-limited, with a resolution defined by the modulation transfer function dropping to half for features with pitch $6\ \mu\text{m}$ in the object plane. Sufficient magnification was used so that the finite pixel-size of our camera caused no degradation of the image.

The image sequence began with 24 images taken of the same atomic sample at a strobe frequency of $10\ \text{kHz}$ while the magnetic field remained oriented along the $\hat{z}$ direction. These images were used to measure the orientation and magnitude of the transverse magnetization. The normalization constant $\zeta(\rho)$ was obtained by averaging the signal over all image frames and fitting with a Thomas–Fermi distribution for the density of the condensate. Following this sequence, two additional image frames were obtained within $5\ \text{ms}$ in which the magnetic field was adiabatically reoriented in the $\hat{y}$ and $-\hat{y}$ directions. The longitudinal magnetization was determined from the difference between these last frames.

Image analysis. The identification of spin vortices in the magnetization images relied on the identification of a core of ‘dark’ pixels, consistent with zero Larmor precession amplitude, which were surrounded by ‘bright’ pixels with a finite amplitude and well-defined phase of Larmor precession. We identified vortices based on two criteria: (1) that there be an island of dark pixels, a candidate for the vortex core, which is surrounded entirely by bright pixels and that is at least two bright pixels away from nearby dark pixels, and (2) that the transverse magnetization traced along a closed loop through bright pixels surrounding

the core have a non-zero net winding. The distinction between bright and dark pixels (at about one-quarter of the maximum Larmor precession amplitude) was chosen to eliminate false-positive vortex detections in simulated data taking into account the measured noise (optical shot-noise limited) and resolution of our imaging system, and using estimates for the width of contiguous domain walls.

Received 23 March; accepted 20 July 2006.

1. Stenger, J. *et al.* Spin domains in ground state spinor Bose–Einstein condensates. *Nature* **396**, 345–348 (1998).
2. Ho, T.-L. Spinor Bose condensates in optical traps. *Phys. Rev. Lett.* **81**, 742–745 (1998).
3. Ohmi, T. & Machida, K. Bose–Einstein condensation with internal degrees of freedom in alkali atom gases. *J. Phys. Soc. Jpn* **67**, 1822–1825 (1998).
4. Saito, H., Kawaguchi, Y. & Ueda, M. Breaking of chiral symmetry and spontaneous rotation in a spinor Bose–Einstein condensate. *Phys. Rev. Lett.* **96**, 065302 (2006).
5. Chang, M.-S. *et al.* Observation of spinor dynamics in optically trapped Rb Bose–Einstein condensates. *Phys. Rev. Lett.* **92**, 140403 (2004).
6. Schmaljohann, H. *et al.* Dynamics of $F = 2$ spinor Bose–Einstein condensates. *Phys. Rev. Lett.* **92**, 040402 (2004).
7. Klausen, N. N., Bohn, J. L. & Greene, C. H. Nature of spinor Bose–Einstein condensates in rubidium. *Phys. Rev. A* **64**, 053602 (2001).
8. Higbie, J. M. *et al.* Direct, non-destructive imaging of magnetization in a spin-1 Bose gas. *Phys. Rev. Lett.* **95**, 050401 (2005).
9. Pu, H. *et al.* Spin-mixing dynamics of a spinor Bose–Einstein condensate. *Phys. Rev. A* **60**, 1463–1470 (1999).
10. Robins, N. P., Zhang, W., Ostrovskaya, E. A. & Kivshar, Y. S. Modulational instability of spinor condensates. *Phys. Rev. A* **64**, 021601(R) (2001).
11. Saito, H. & Ueda, M. Spontaneous magnetization and structure formation in a spin-1 ferromagnetic Bose–Einstein condensate. *Phys. Rev. A* **72**, 023610 (2005).
12. Zhang, W. *et al.* Dynamical instability and domain formation in a spin-1 Bose–Einstein condensate. *Phys. Rev. Lett.* **95**, 180403 (2005).
13. Widera, A. *et al.* Coherent collisional spin dynamics in optical lattices. *Phys. Rev. Lett.* **95**, 190405 (2005).
14. Chang, M.-S. *et al.* Coherent spinor dynamics in a spin-1 Bose condensate. *Nature Phys.* **1**, 111–116 (2005).
15. Hall, D. S. *et al.* The dynamics of component separation in a binary mixture of Bose–Einstein condensates. *Phys. Rev. Lett.* **81**, 1539–1542 (1998).
16. Miesner, H.-J. *et al.* Observation of metastable states in spinor Bose–Einstein condensates. *Phys. Rev. Lett.* **82**, 2228–2231 (1999).
17. Stamper-Kurn, D. M. & Ketterle, W. in *Coherent Matter Waves* (eds Kaiser, R., Westbrook, C. & David, F.) 137–218 (Springer, New York, 2001).
18. Goldstein, E. V. & Meystre, P. Quasiparticle instabilities in multicomponent atomic condensates. *Phys. Rev. A* **55**, 2935–2940 (1997).
19. Timmermans, E. Phase separation of Bose–Einstein condensates. *Phys. Rev. Lett.* **81**, 5718–5721 (1997).
20. Kibble, T. W. B. Topology of cosmic domains and strings. *J. Phys. A* **9**, 1387–1398 (1976).
21. Zurek, W. H. Cosmological experiments in superfluid helium? *Nature* **317**, 505–508 (1985).
22. Zurek, W. H., Dornier, U. & Zoller, P. Dynamics of a quantum phase transition. *Phys. Rev. Lett.* **95**, 105701 (2005).
23. Chuang, I., Durrer, R., Turok, N. & Yurke, B. Cosmology in the laboratory: defect dynamics in liquid crystals. *Science* **251**, 1336–1342 (1991).
24. Hendry, P. C. *et al.* Generation of defects in superfluid ^4He as an analogue of the formation of cosmic strings. *Nature* **368**, 315–317 (1994).
25. Ruutu, V. M. H. *et al.* Vortex formation in neutron-irradiated superfluid He-3 as an analogue of cosmological defect formation. *Nature* **382**, 334–336 (1996).
26. Bauerle, C. *et al.* Laboratory simulation of cosmic string formation in the early Universe using superfluid He-3. *Nature* **382**, 332–334 (1996).
27. Isoshima, T., Machida, K. & Ohmi, T. Quantum vortex in a spinor Bose–Einstein condensate. *J. Phys. Soc. Jpn* **70**, 1604–1610 (2001).
28. Mermin, N. D. & Ho, T.-L. Circulation and angular momentum in the A phase of superfluid helium-3. *Phys. Rev. Lett.* **36**, 594–597 (1976).

Acknowledgements We thank E. Mueller, J. Moore, and A. Vishwanath for comments, J. Guzman for experimental assistance, and the NSF and the David and Lucile Packard Foundation for financial support. S.R.L. acknowledges support from the NSERC.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.M.S.-K. (dmsk@berkeley.edu).

LETTERS

'Designer atoms' for quantum metrology

C. F. Roos^{1,2}, M. Chwalla¹, K. Kim¹, M. Riebe¹ & R. Blatt^{1,2}

Entanglement is recognized as a key resource for quantum computation¹ and quantum cryptography². For quantum metrology, the use of entangled states has been discussed^{3–5} and demonstrated⁶ as a means of improving the signal-to-noise ratio. In addition, entangled states have been used in experiments for efficient quantum state detection⁷ and for the measurement of scattering lengths⁸. In quantum information processing, manipulation of individual quantum bits allows for the tailored design of specific states that are insensitive to the detrimental influences of an environment⁹. Such 'decoherence-free subspaces' (ref. 10) protect quantum information and yield significantly enhanced coherence times¹¹. Here we use a decoherence-free subspace with specifically designed entangled states¹² to demonstrate precision spectroscopy of a pair of trapped Ca⁺ ions; we obtain the electric quadrupole moment, which is of use for frequency standard applications. We find that entangled states are not only useful for enhancing the signal-to-noise ratio in frequency measurements—a suitably designed pair of atoms also allows clock measurements in the presence of strong technical noise. Our technique makes explicit use of non-locality as an entanglement property and provides an approach for 'designed' quantum metrology.

Metrology seeks to relate all measurements to a precision determination of frequencies. The most precise measurements today are obtained from measuring atomic transition frequencies by monitoring the phase evolution of a superposition of the pertaining states. This is achieved by applying Ramsey's interferometric technique¹³ that can be generalized⁵ to (maximally) entangled states: under free precession, the two-atom state $\Psi = \frac{1}{\sqrt{2}}(|u_1\rangle|u_2\rangle + |v_1\rangle|v_2\rangle)$ evolves into the state $\Psi(\tau) = \frac{1}{\sqrt{2}}(|u_1\rangle|u_2\rangle + \exp(i\lambda_\phi\tau)|v_1\rangle|v_2\rangle)$ where the phase evolution rate $\lambda_\phi = [(E_{u_1} + E_{u_2}) - (E_{v_1} + E_{v_2})]/\hbar$ is proportional to the differences in atomic energies E_{u_k}, E_{v_k} of the involved atomic levels u_k, v_k . (Here τ is the free precession time.) The real part of the phase factor $e^{i\lambda_\phi\tau}$ is measured by projecting the atoms onto the states $|\pm\rangle_k = \frac{1}{\sqrt{2}}(|u_k\rangle \pm |v_k\rangle)$ and measuring the parity operator $P = \hat{P}_{(++)} + \hat{P}_{(--)}$ — $\hat{P}_{(+-)} - \hat{P}_{(-+)}$, where $\hat{P}_{(\pm\pm)}$ denotes the projector onto the state $|\pm\rangle_1|\pm\rangle_2$. If Ψ belongs to a decoherence-free subspace, values of τ of several seconds^{14,15} allow for highly accurate phase estimation.

In atomic optical frequency standards based on single hydrogen-like trapped ions¹⁶, the transition frequency from the *S* ground state to a metastable *D_j* state is measured by an optical frequency comb¹⁷. The *D_j* state's atomic electric quadrupole moment interacting with residual electric quadrupole fields¹⁸ gives rise to frequency shifts of a few hertz. The shift of the Zeeman sublevel $|D_j, m_j\rangle$ in a quadrupole field $\Phi(x, y, z) = A(x^2 + y^2 - 2z^2)$ is given by:

$$\hbar\Delta\nu = \frac{1}{4} \frac{dE_z}{dz} \Theta(D, j) \frac{j(j+1) - 3m_j^2}{j(2j-1)} (3\cos^2\beta - 1) \quad (1)$$

where $dE_z/dz = 4A$ is the electric field gradient along the potential's symmetry axis *z*, β denotes the angle between the quantization axis

and *z*, and where $\Theta(D, j)$ expresses the strength of the quadrupole moment in terms of a reduced matrix element¹⁸.

Recently, quadrupole moments have been measured for ⁸⁸Sr⁺, ¹⁹⁹Hg⁺ and ¹⁷¹Yb⁺ with a precision ranging from about 4% to 12% (refs 19–21). In these single-ion experiments, narrow-linewidth lasers are used to detect electric quadrupole shifts by measuring the transition frequency from the electronic ground state to the metastable state. Alternatively, the quadrupole shift could also be measured by performing Ramsey experiments with an ion prepared in a superposition of Zeeman *D_j* sublevels. Both measurement schemes are subject to phase decoherence. Laser frequency noise limits the coherence time if the *S* state is used as the reference state. Magnetic field noise giving rise to first-order Zeeman shifts renders the measurement scheme based on *D* state superpositions difficult. However, both sources of phase noise can be eliminated by replacing the single ion by a two-ion entangled state prepared in a decoherence-free subspace. Ramsey experiments with the two-ion Bell state $\Psi = \frac{1}{\sqrt{2}}(|m_1\rangle|m_2\rangle + |m_3\rangle|m_4\rangle)$, where the magnetic quantum numbers m_i of the states $|m_i\rangle \equiv |D_j, m_i\rangle$ satisfy $m_1 + m_2 = m_3 + m_4$, are not affected by fluctuations of the magnetic fields to first order because both parts of the superposition are Zeeman-shifted by the same amount (Fig. 1b). Also, frequency noise of the laser used for preparing Ψ is relevant only during the comparatively short state-preparation and read-out steps but not during the long interrogation period. For suitably chosen values of $m_1, \dots, m_4$, a decoherence-immune state can be designed that is sensitive to the electric quadrupole shift (Fig. 1c).

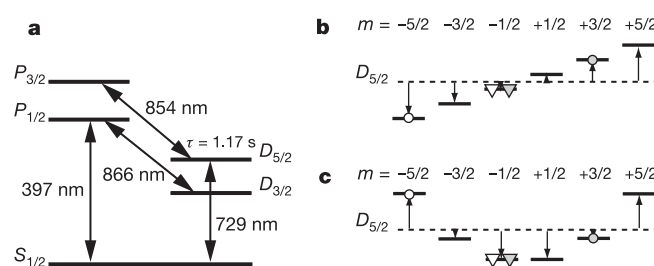


Figure 1 | Relevant atomic levels of ⁴⁰Ca⁺. **a**, Bell states are prepared by coherently exciting the *S*_{1/2} ↔ *D*_{5/2} quadrupole transition. The *S*_{1/2} ↔ *P*_{1/2} transition is used for quantum state read-out, and lasers at 866 and 854 nm wavelength serve as repumping and quenching lasers. **b**, Energy level shift of the $|m\rangle \equiv |D_{5/2}, m\rangle$ states in a weak magnetic field; **c**, energy level shift caused by an electric quadrupole potential. Energy levels occurring in the state $\Psi_1 = \frac{1}{\sqrt{2}}(|o\rangle|\bullet\rangle + |\nabla\rangle|\nabla\rangle)$ (equation (2)) are denoted by the symbols $\circ, \nabla, \bullet, \nabla$, the open (or filled) symbols corresponding to levels occupied by atom 1 (or 2), respectively. The combined Zeeman energy of the $|o\rangle|\bullet\rangle$ states equals the energy of the $|\nabla\rangle|\nabla\rangle$ states, making Ψ_1 immune against magnetic field noise. As an electric quadrupole field shifts the $|o\rangle|\bullet\rangle$ and $|\nabla\rangle|\nabla\rangle$ states in opposite directions, Ψ_1 is suitable for a measurement of the quadrupole shift.

¹Institut für Experimentalphysik, Universität Innsbruck, Technikerstr. 25, A-6020 Innsbruck, Austria. ²Institut für Quantenoptik und Quanteninformation, Österreichische Akademie der Wissenschaften, Otto-Hittmair-Platz 1, A-6020 Innsbruck, Austria.

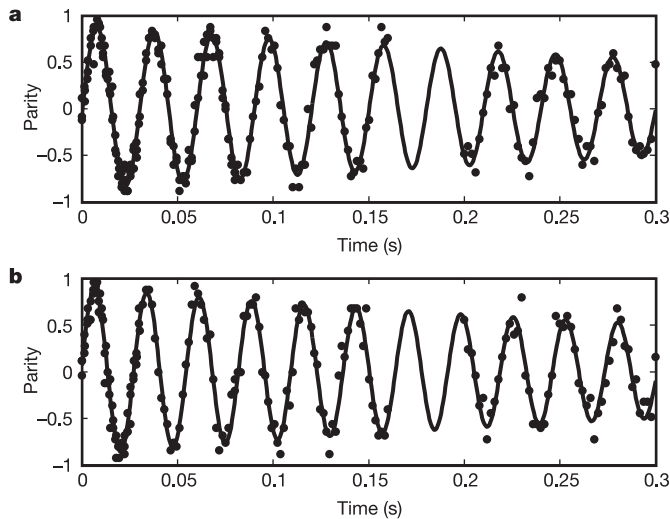


Figure 2 | Parity oscillations of the entangled states Ψ_1 and Ψ_2 at $U_{\text{tips}} = 750$ V tip voltage. The experiment is repeated 100 times for each data point. An exponentially damped sinusoidal function is fitted to the data. **a**, For Ψ_1 , the oscillation frequency is $\Delta_1 = (2\pi) 33.35(3)$ Hz and the damping time 587(70) ms. **b**, For Ψ_2 , the fit yields $\Delta_2 = (2\pi) 36.52(4)$ Hz and a decay time 525(60) ms. For each plot, five different data sets were merged. No data were taken covering waiting times around 170 ms. To exclude slow drifts of the initial parity overtime, data for waiting times up to 20 ms were repeatedly taken.

The isotope $^{40}\text{Ca}^+$ has a single valence electron and no hyperfine structure (Fig. 1a). For its metastable $3d^2D_{5/2}$ state (lifetime $\tau_{D_{5/2}} = 1.168(7)$ s), a quadrupole moment of $1.917ea_0^2$ was calculated^{22,23} (here e is the electronic charge and a_0 is the Bohr radius). (All numbers in parentheses denote 1σ confidence intervals for fitted parameters.) For the experiment, two $^{40}\text{Ca}^+$ ions are trapped along the axis of a linear ion trap. By applying a static (d.c.) voltage ranging from 500 V to 2,000 V to the tip electrodes²⁴, axial centre-of-mass (COM) mode trap frequencies ω_z ranging from 850 kHz to 1,700 kHz are achieved, leading to distances between the ions from $6.2\ \mu\text{m}$ to $3.9\ \mu\text{m}$. A charge-coupled device (CCD) camera images the ions' fluorescence. The degeneracy of the Zeeman states is lifted by applying a magnetic bias field of 2.9 G. The ions are cooled to the ground state of the axial COM mode by Doppler and sideband cooling²⁵. For coherent quantum state manipulation, the ions are individually excited on the $|S\rangle \equiv S_{1/2}(m = -1/2) \leftrightarrow |D\rangle \equiv D_{5/2}(m = -1/2)$ transition by a tightly focused laser beam (laser linewidth ≈ 150 Hz) resonant with either the carrier transition or the upper motional sideband. For a detailed description of the experimental set-up, see ref. 24.

The Bell state $\Psi_{SD}^{\pm} = (|S\rangle|D\rangle \pm |D\rangle|S\rangle)/\sqrt{2}$ is created with a fidelity of about 90% by a sequence of three laser pulses¹¹. Additional carrier π pulses transfer the entanglement into the $D_{5/2}$ Zeeman state manifold, thus generating the state:

$$\Psi_1 = \frac{1}{\sqrt{2}}(|-5/2\rangle|+3/2\rangle + |-1/2\rangle|-1/2\rangle) \quad (2)$$

This magnetic-field-insensitive state is used for a measurement of the quadrupole shift. In the vicinity of the trap centre, the d.c. voltage applied to the tip electrodes creates a rotationally symmetric electric quadrupole potential, which shifts the energy of the constituents of the superposition state Ψ_1 by $\hbar\Delta_1 = 24/5 \hbar\delta$ with respect to each other. Here, $\hbar\delta$ is the shift that a single ion in state $|-5/2\rangle$ would experience. The two-ion shift is bigger by $12/5$ because two ions contribute that are prepared in substates that shift in opposite directions. In addition, the presence of a second ion doubles the electric field gradient at the location of the other ion¹².

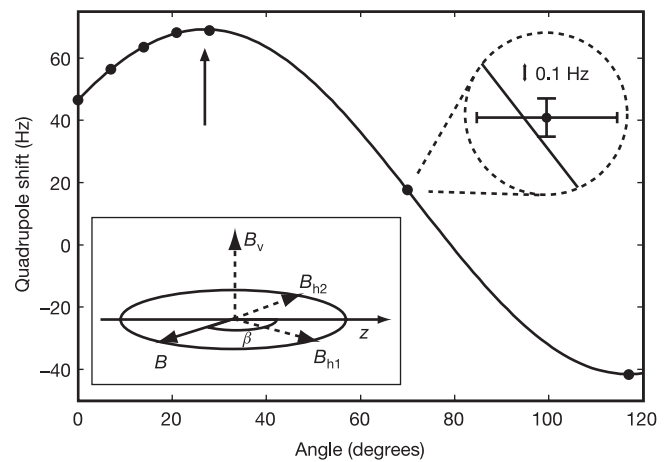


Figure 3 | Angular dependence of the quadrupole shift. Main panel, quadrupole shift $\Delta/(2\pi)$ at $U_{\text{tips}} = 1,000$ V as a function of the magnetic field orientation $\beta - \beta_0$. The data are fitted by $\Delta = \Delta_a + \Delta_b \cos^2(\beta - \beta_0)$, yielding $\beta_0 = 26.9^\circ$. The arrow indicates the angle corresponding to $\beta = 0$. For the determination of the quadrupole moment, shift measurements are performed at $\beta = 0$ as errors in β enter the calculation only in second order. Note that the error bars are smaller than the point size. Upper right inset, an enlarged data point. Error bars, 1σ . Lower left inset, the geometry of the magnetic field coils. The angle β is varied by appropriately changing the current in pairs of coils producing the magnetic fields B_{h1} , B_{h2} . Field B_v is used to null residual fields in the perpendicular direction.

For the shift measurement, Ψ_1 is allowed to evolve into $\Psi_1(\tau) = \frac{1}{\sqrt{2}}(|-5/2\rangle|+3/2\rangle + \exp(i\Delta_1\tau)|-1/2\rangle|-1/2\rangle)$ and $\cos(\Delta_1\tau)$ is measured for τ ranging from 0 ms to 300 ms. Figure 2a shows the resulting parity oscillations¹¹ at a frequency $\Delta_1 = (2\pi) 33.35(3)$ Hz. Disentanglement of $\Psi_1(\tau)$ by spontaneous decay manifests itself as damping of the oscillations with a damping time constant $\tau_d = 587(70)$ ms, close to the expected value $\frac{1}{2}\tau_{D_{5/2}} = 584$ ms. Parity oscillations of the state $\Psi_2(\tau) = \frac{1}{\sqrt{2}}(|+3/2\rangle|-5/2\rangle + \exp(i\Delta_2\tau)|-1/2\rangle|-1/2\rangle)$ are shown in Fig. 2b. Its oscillation frequency $\Delta_2 = (2\pi) 36.52(4)$ Hz differs slightly from Δ_1 because of a magnetic field gradient in the direction of the ion crystal that gives rise to an additional contribution to the parity signal¹². Both phase oscillation contributions can be separately determined by taking the average $\Delta = (\Delta_1 + \Delta_2)/2$ and the difference $\Delta_B = |\Delta_1 - \Delta_2|/2$ of the signals. By recording parity oscillations at different tip voltages U_{tips} , we measure the quadrupole shift $\Delta/(2\pi)$ as a function of the electric field gradient and observe a linear change in $\Delta(U_{\text{tips}})$.

For a precise determination of the quadrupole moment, the dependence of the quadrupole shift on the orientation of the magnetic field should be minimized. We investigate the angle dependence of Δ by varying the orientation of the magnetic field in a plane that also contains the z axis of the trap. Initialization of the ions in a pure state before coherent manipulation is achieved by optical pumping on the quadrupole transition (see Methods). Generally, it cannot be excluded that the quadrupole field has imperfect rotational symmetry. In this case, the factor $3\cos^2\beta - 1$ in equation (1) needs to be replaced by $(3\cos^2\beta - 1) - \epsilon \sin^2\beta \cos(2\alpha)$, where ϵ characterizes the asymmetry and α its direction¹⁸. Note, however, that the additional term vanishes for $\beta = 0$. Figure 3 shows the sinusoidal variation of Δ measured at $U_{\text{tips}} = 1,000$ V as a function of the angle. At the angle that maximizes Δ , another maximization is performed in the perpendicular direction to determine the direction corresponding to $\beta = 0$. Figure 4a displays the quadrupole shift measured at $\beta = 0$ as a function of the electric field gradient dE_z/dz . The influence of the magnetic field gradient on the phase evolution rate is shown in Fig. 4b. Using $dE_z/dz = -m\omega_z^2/e$, the gradient is calibrated by a measurement of the ion's oscillation frequency ω_z . By fitting a straight line $\Delta/(2\pi) = \Delta_0/(2\pi) + a(dE_z/dz)$ to the data, we

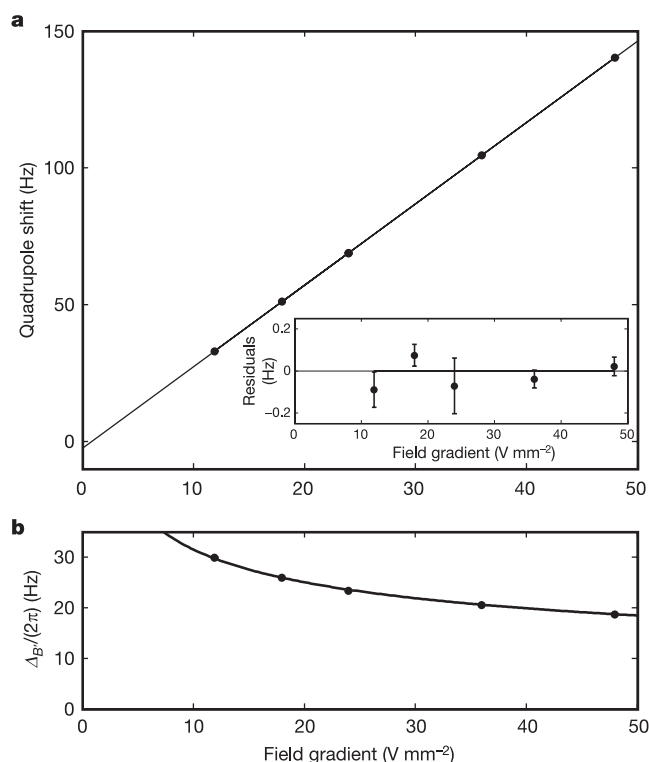


Figure 4 | Level shifts as a function of the applied external electric field gradient, dE_z/dz . **a**, Quadrupole shift. The offset at $dE_z/dz = 0$ is mainly due to the second-order Zeeman effect. The quadrupole moment is proportional to the slope, which is measured with an uncertainty of less than 0.1%. Inset, the deviation of the data points from the linear fit. Error bars, 1σ . **b**, Phase oscillation rate $\Delta_B/(2\pi)$ due to the magnetic field gradient calculated from $|\Delta_2 - \Delta_1|/2$. As the distance d between the ions decreases at higher electric field gradients, $\Delta_B \propto d \propto (dE_z/dz)^{-1/3}$.

determine the slope $a = 2.975(2) \text{ Hz mm}^2 \text{ V}^{-1}$ and the offset $\Delta_0/(2\pi) = -2.4(1) \text{ Hz}$. The major part of the phase oscillation frequency Δ_0 at $dE_z/dz = 0$ is caused by the second-order Zeeman effect that contributes $\Delta_{B^2}/(2\pi) = -2.9 \text{ Hz}$ at the bias field of 2.9 G. The remaining part $\Delta_0 - \Delta_{B^2}$ is attributed to a residual quadrupole field caused by stray charges; a.c.-Stark shifts are insignificant. The quadrupole moment $\Theta(3d, 5/2) = \frac{5}{12} \hbar a$ is proportional to the slope a . Whereas the uncertainty in the value of a is smaller than 10^{-3} , the main uncertainty in $\Theta(3d, 5/2)$ is due to errors in the determination of the angle β . Assuming that the angle can be determined with an accuracy of 3° (see Methods), we determine the quadrupole moment to be $\Theta(3d, 5/2) = 1.83(1)ea_0^2$.

These results show the viability of the approach of designing specific entangled states for precision measurements. For frequency-standard applications, minimization of the quadrupole shift is imperative^{26,27}. Note that, by using designed entangled states, the quadrupole shift could be cancelled by averaging the results over states with different magnetic quantum numbers¹². The design approach demonstrated here is new to metrology: whereas previously the use of entangled states was to enhance the signal-to-noise ratio, the use of a designed pair of entangled ions allows an atomic clock measurement to be performed in a subspace free of magnetic field noise. This is achieved with unprecedented precision on an even isotope (zero nuclear spin), which opens the way to optical frequency standards based on a whole variety of easily accessible atomic systems not considered so far.

METHODS

Frequency-resolved optical pumping. Owing to the magnetic bias field, the ions can be prepared in the $|S_{1/2}, m = -1/2\rangle$ state for an arbitrary orientation of the magnetic field by frequency-resolved optical pumping on the quadrupole

transition. For this, the $|S_{1/2}, m = +1/2\rangle \leftrightarrow |D_{5/2}, m = -3/2\rangle$ transition is excited while the repumping lasers on the $D_{5/2} \leftrightarrow P_{3/2}$ and $D_{3/2} \leftrightarrow P_{1/2}$ transitions are switched on at the same time. After an excitation of 1 ms duration, at least 99.9% of the population has been pumped to the $|S_{1/2}, m = -1/2\rangle$ state.

Orientation of the magnetic field. The magnetic field $\mathbf{B}$ is set by passing currents I_{h_1}, I_{h_2}, I_v through three mutually orthogonal pairs of coils (Fig. 3 lower left inset). We calibrate the field coils and null the offset fields by measuring the Zeeman shift of the $|S_{1/2}, m = -1/2\rangle \leftrightarrow |D_{5/2}, m = -1/2\rangle$ transition as a function of the currents. We estimate that the offset fields can be zeroed to a value below 1% of the applied bias field. After nulling the field in the v -direction, we map out the values of (I_{h_1}, I_{h_2}) that yield a constant Zeeman shift and fit an ellipse to the data. These current values rotate $\mathbf{B}$ in a horizontal plane that also contains the symmetry axis (z) of the ion trap. A similar procedure is applied for calibrating the v -coil.

Alignment of the magnetic field with the trap axis. We align the direction of the magnetic field with the trap axis by maximizing the quadrupole shift as a function of the field direction. A residual quadrupole potential Φ_s caused by stray charges and by voltages applied to the compensation electrodes will in general have principal axes that do not coincide with those of the trap's quadrupole potential. This will lead to errors in the alignment of the magnetic field $\mathbf{B}$ with the principal axis of the trap's electric quadrupole potential Φ_{tips} . This is due to the fact that $\mathbf{B}$ is aligned with the direction $\mathbf{n}_{\text{ts}}$ that maximizes the quadrupole shift of $\Phi_{\text{tips}}(\mathbf{n}) + \Phi_s(\mathbf{n})$ instead of the direction $\mathbf{n}_t$ that would be chosen without Φ_s . Fortunately, Φ_s is much smaller than Φ_{tips} in the experiment. We obtain information about the magnitude of Φ_s by: (1) extrapolating the ion oscillation frequency $\omega_z(U_{\text{tips}})$ for $U_{\text{tips}} \rightarrow 0$; (2) measuring the electric stray field $\mathbf{E}_s$ and assuming the magnitude of the field gradient to be $\propto |\mathbf{E}_s|/r$, where r is the smallest distance of the ions from the trap electrodes; and (3) most importantly, by measuring the residual quadrupole shifts for $\beta = 0$ and another direction with $\beta = 26.9^\circ$, constraints regarding the shape, orientation and magnitude of Φ_s can be found that allow us to calculate the average angle $\Delta\beta$ between $\mathbf{n}_{\text{ts}}$ and $\mathbf{n}_t$. We find $\Delta\beta = 3^\circ$, and use this value for estimating the uncertainty in the determination of $\Theta(3d, 5/2)$.

Received 5 May; accepted 18 July 2006.

- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, Cambridge, UK, 2000).
- Gisin, N., Ribordy, G., Tittel, W. & Zbinden, H. Quantum cryptography. *Rev. Mod. Phys.* **74**, 145–195 (2002).
- Giovannetti, V., Lloyd, S. & Maccone, L. Quantum-enhanced measurements: beating the standard quantum limit. *Science* **306**, 1330–1336 (2004).
- Giovannetti, V., Lloyd, S. & Maccone, L. Quantum metrology. *Phys. Rev. Lett.* **96**, 010401 (2006).
- Bollinger, J. J., Itano, W. M., Wineland, D. J. & Heinzen, D. J. Optimal frequency measurements with maximally correlated states. *Phys. Rev. A* **54**, R4649–R4652 (1996).
- Leibfried, D. et al. Toward Heisenberg-limited spectroscopy with multiparticle entangled states. *Science* **304**, 1476–1478 (2004).
- Schmidt, P. O. et al. Spectroscopy using quantum logic. *Science* **309**, 749–752 (2005).
- Widera, A. et al. Entanglement interferometry for precision measurement of atomic scattering properties. *Phys. Rev. Lett.* **92**, 160406 (2004).
- Kielinski, D. et al. A decoherence-free quantum memory using trapped ions. *Science* **291**, 1013–1015 (2001).
- Lidar, D. A., Chuang, I. L. & Whaley, K. B. Decoherence-free subspaces for quantum computation. *Phys. Rev. Lett.* **81**, 2594–2597 (1998).
- Roos, C. F. et al. Bell states with ultra-long lifetimes and their tomographic state analysis. *Phys. Rev. Lett.* **92**, 220402 (2004).
- Roos, C. F. Precision frequency measurements with entangled states. Preprint at (<http://arxiv.org/quant-ph/0508148>) (2005).
- Ramsey, N. F. A molecular beam resonance method with separated oscillating fields. *Phys. Rev.* **78**, 695–699 (1950).
- Häffner, H. et al. Robust entanglement. *Appl. Phys. B* **81**, 151–153 (2005).
- Langer, C. et al. Long-lived qubit memory using atomic ions. *Phys. Rev. Lett.* **95**, 060502 (2005).
- Madej, A. A. & Bernard, J. E. *Frequency Measurement and Control* (ed. Luiten, A. N.) 153–195 (Topics in Applied Physics Vol. 79, Springer, Berlin/Heidelberg, 2001).
- Diddams, S. A., Bergquist, J. C., Jefferts, S. R. & Oates, C. W. Standards of time and frequency at the outset of the 21st century. *Science* **306**, 1318–1324 (2004).
- Itano, W. M. External-field shifts of the $^{199}\text{Hg}^+$ optical frequency standard. *J. Res. Natl. Inst. Stand. Technol.* **105**, 829–837 (2000).
- Oskey, W. H., Itano, W. M. & Bergquist, J. C. Measurements of the $^{199}\text{Hg}^+ 5d^9 6s^2 \ ^2D_{5/2}$ electric quadrupole moment and a constraint on the quadrupole shift. *Phys. Rev. Lett.* **94**, 163001 (2005).

20. Barwood, G. P., Margolis, H. S., Huang, G., Gill, P. & Klein, H. A. Measurement of the electric quadrupole moment of the $4d\ ^2D_{5/2}$ level in $^{88}\text{Sr}^+$. *Phys. Rev. Lett.* **93**, 133001 (2004).
21. Schneider, T., Peik, E. & Tamm, Chr. Sub-Hertz optical frequency comparisons between two trapped $^{171}\text{Yb}^+$ ions. *Phys. Rev. Lett.* **94**, 230801 (2005).
22. Itano, W. M. Quadrupole moments and hyperfine constants of metastable states of Ca^+ , Sr^+ , Ba^+ , Yb^+ , Hg^+ , and Au . *Phys. Rev. A* **73**, 022510 (2006).
23. Sur, C. *et al.* Electric quadrupole moments of the D states of alkaline-earth-metal ions. *Phys. Rev. Lett.* **96**, 193001 (2006).
24. Schmidt-Kaler, F. *et al.* How to realize a universal quantum gate with trapped ions. *Appl. Phys. B* **77**, 789–796 (2003).
25. Roos, Ch. *et al.* Quantum state engineering on an optical transition and decoherence in a Paul trap. *Phys. Rev. Lett.* **83**, 4713–4716 (1999).
26. Margolis, H. S. *et al.* Hertz-level measurement of the optical clock frequency in a single $^{88}\text{Sr}^+$ ion. *Science* **306**, 1355–1358 (2004).
27. Dubé, P. *et al.* Electric quadrupole shift cancellation in single-ion optical frequency standards. *Phys. Rev. Lett.* **95**, 033001 (2005).

Acknowledgements We thank H. Häffner for his contributions to the optical pumping scheme and acknowledge help with the experiment from T. Körber, W. Hänsel, D. Chek-al-kar, M. Mukherjee and P. Schmidt. We acknowledge support by the Austrian Science Fund (FWF), by the European Commission (SCALA, CONQUEST networks), and by the Institut für Quanteninformation GmbH. This material is based on work supported in part by the US Army Research Office.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.F.R. (christian.roos@uibk.ac.at).

LETTERS

A guest-free germanium clathrate

Arnold M. Guloy^{1,2}, Reiner Ramlau¹, Zhongjia Tang^{1,2}, Walter Schnelle¹, Michael Baitinger¹ & Yuri Grin¹

The challenges associated with synthesizing expanded semiconductor frameworks with cage-like crystal structures continue to be of interest^{1,2}. Filled low-density germanium and silicon framework structures have distinct properties that address important issues in thermoelectric phonon glass–electron crystals³, superconductivity⁴ and the possibility of Kondo insulators⁵. Interest in empty framework structures of silicon and germanium is motivated by their predicted wide optical bandgaps of the same magnitude as quantum dots and porous silicon, making them and their alloys promising materials for silicon-based optoelectronic devices^{6,7}. Although almost-empty $\text{Na}_{1-x}\text{Si}_{136}$ has already been reported^{8,9}, the synthesis of guest-free germanium clathrate has so far been unsuccessful. Here we report the high-yield synthesis and characteristics of germanium with the empty clathrate-II structure through the oxidation of Ge_9^{4-} Zintl anions in ionic liquids under ambient conditions. The approach demonstrates the potential of ionic liquids as media for the reactions of polar intermetallic phases.

The advent of low-temperature synthesis has brought new opportunities in the field of materials synthesis and has provided new routes to inorganic solids that are usually prepared by high-temperature reactions. The use of soft-chemical synthetic methods is common in the preparation of a wide range of zeolites and chalcogenides^{10,11}. However, application of soft-chemical techniques to less ionic materials such as intermetallic phases of group IV elements is still relatively uncharted, although potentially of great importance as a source of unique electronic materials. Success has been limited to topotactic reactions, as in the formation of *allo*-Ge and *allo*-Si (of as yet unknown structure) by the removal of the cations from the layered Zintl phases with layered crystal structures, $\text{Li}_{17}\text{Ge}_{12}$ and Li_3NaSi_6 , respectively^{12,13}.

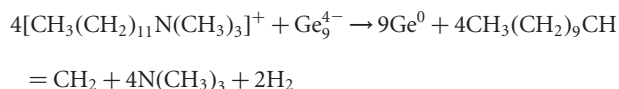
The preparation of empty clathrate-II $\square_{24}\text{Si}_{136}$ ($\square$ means a vacancy on a possible metal position) involves the thermal decomposition of the Zintl-phase NaSi under vacuum, which results in a mixture of clathrates $\text{Na}_x\text{Si}_{136}$ ($x < 1$) and silicon. Separation of $\square_{24}\text{Si}_{136}$ from the heterogeneous product $\text{Na}_{1-x}\text{Si}_{136}$ was achieved by repeated density separation and centrifugation or reaction with iodine vapour. In contrast, there have been no reports of any successful synthesis of an empty or almost empty clathrate $\square_{24}\text{Ge}_{136}$. Theoretical studies have predicted a relatively high stability of $\square_{24}\text{Si}_{136}$ and $\square_{24}\text{Ge}_{136}$ with respect to the diamond form of the elements, and various synthetic methods have been proposed including epitaxial growth on seed crystals, vapour-phase synthesis, and crystallization under ‘strained or stretched’ conditions^{14–16}.

A general technique for preparing materials with low-density crystal structures is inorganic–organic template assembly in the solution phase^{10,17}. This approach exploits favourable electrostatic interactions between organic and inorganic species to assemble soluble inorganic units, with organic species acting as templates. The inorganic units are subsequently condensed into a relatively rigid network, and template agents are removed to produce a low-density

framework solid. Critical to the ‘bottom-up’ hybrid assembly approach is the need for soluble inorganic building blocks, organic templates and a compatible solvent–reaction medium. Typical Zintl anions of Si and Ge such as Si_9^{4-} and Ge_9^{4-} occur as molecular anions in solid-state compounds and present soluble building blocks of technologically important elements^{18,19}. Chemical studies on alkali-metal Zintl phases, containing Ge_9^{4-} clusters, show a stepwise oxidative coupling of Ge_9^{n-} ($n = 2–4$) anions in amine solvents that culminates in the formation of linear polymers of vertex-linked clusters $[\text{Ge}_9^{2-}]_\infty$ (refs 20, 21). Thus, we considered further oxidation of the anionic polymers as a promising route to new materials with the oxidation state of Ge close to zero.

Ionic liquids formed by low-melting AlCl_4^- and PF_6^- -based organic salts and their mixtures have unique properties, such as negligible vapour pressure, low viscosity, high thermal, chemical and electrochemical stability, and differentiated miscibility with organic compounds, making them potentially useful solvents for inorganic synthesis, particularly for the assembly and polymerization of inorganic units into low-density solids²². Reactions in an ionic liquid (IL) also offer ease in separation and characterization, and possible recycling of the solvent. Quaternary alkyltrimethylammonium-based ILs also have the advantage of being surfactants that can act as a template for the growth of porous structures, as shown in the preparation of microporous oxide materials such as aluminophosphate zeolites²³.

The synthesis of crystalline $\square_{24}\text{Ge}_{136}$ under ambient conditions can be rationalized by the formation of a three-dimensional network of four-bonded germanium atoms (Ge^0) from the polymerization and oxidation of Ge_9^{4-} anions. By taking advantage of the inherent reactivity of Na_4Ge_9 in IL, soft oxidation of Ge_9^{4-} was performed at 300 °C in a eutectic mixture of dodecyltrimethylammonium chloride (DTAC) and aluminium trichloride (1:1 molar ratio). The high-yield redox reaction between the Ge_9 anions and DTAC follows the Hofmann elimination mechanism²⁴, as demonstrated by the formation of trimethylamine and a paraffin layer (1-dodecene) at the top of the resulting solidified reaction mixtures, and results in the formation of a grey polycrystalline product (see Supplementary Fig. S1):



The low acidity of the β -hydrogen atoms in DTAC permits the controlled oxidation of the Ge_9^{4-} anions to $\square_{24}\text{Ge}_{136}$. The use of more acidic ammonium halides, for example NH_4X ($\text{X} = \text{Cl}, \text{Br}$ or I), instead of DTAC led to the rapid formation of a mixture of amorphous and crystalline α -Ge.

The XRD pattern of the major phase was completely indexed with a face-centred lattice and the cubic lattice parameter of $a = 15.2115(1)$ Å. This is notably smaller than those for $\text{Na}_x\text{Ge}_{136}$ ($a = 15.4$ Å)²⁵, $\text{Cs}_8\text{Ge}_{136}$ ($a = 15.329$ Å)²⁶ and $\text{Cs}_8\text{Na}_{16}\text{Ge}_{136}$

¹Max-Planck-Institut für Chemische Physik fester Stoffe, 01187 Dresden, Germany. ²Department of Chemistry and the Texas Center for Superconductivity, University of Houston, Houston, Texas 77204-5003, USA.

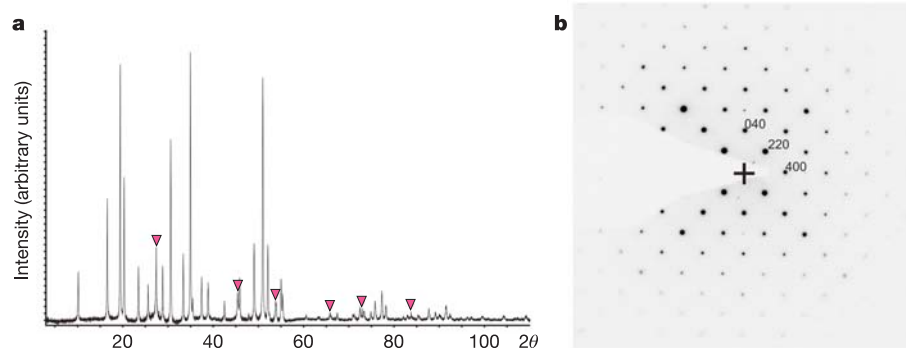


Figure 1 | Crystal structure determination of $\square_{24}\text{Ge}_{136}$. **a**, Experimental XRPD intensities of $\square_{24}\text{Ge}_{136}$ as a function of diffraction angle; reflections from the minor phase $\alpha\text{-Ge}$ are marked. **b**, SAED pattern for the [100] zone axis, confirming the cubic symmetry and cell parameters determined by XRPD.

($a = 15.480(1) \text{ \AA}$)²⁷ and already indicates a low occupation of the cation positions. Aside from the sharp reflections of the clathrate phase, the diffraction patterns showed a significant background, which points to an additional amorphous product. X-ray powder diffraction (XRPD) of the polycrystalline product also indicated the presence of a small amount of $\alpha\text{-Ge}$. Refinement of the crystal structure from powder diffraction data (Fig. 1a) confirmed the clathrate-II framework of $\square_{24}\text{Ge}_{136}$. The cationic positions were found to be empty within the estimated standard deviations (see Supplementary Tables S1–S3).

However, chemical analysis of the complete crystalline and amorphous bulk product, using inductively coupled plasma optical emission spectroscopy, revealed the presence of 0.8% sodium by mass. High-resolution transmission electron microscopy (HRTEM) showed the product to contain one major crystalline phase ($\square_{24}\text{Ge}_{136}$) and two minor phases, one being amorphous and one crystalline. The selected-area electron diffraction (SAED) patterns confirmed the cubic symmetry and lattice parameters of a clathrate-II structure (Fig. 1b). The agreement between experimental and simulated HRTEM images (Fig. 2a, b) was rather good, confirming the results of the X-ray structure refinement. However, the occupancy within the cages of the structure cannot be excluded unequivocally from HRTEM data. Only investigations by electron energy-loss spectroscopy (EELS) revealed the absence of Na from the Ge framework: no Na-*K* edge at 1,072 eV was observed in the EEL spectra of the major crystalline phase. In contrast, a very weak Na-*K* signal appears in the EEL spectra of the amorphous Ge phase (Fig. 3). Thus, the residual Na in the bulk material is contained mainly in the amorphous minor phase, and the cages of crystalline $\square_{24}\text{Ge}_{136}$ are essentially empty.

The magnetic susceptibility of $\square_{24}\text{Ge}_{136}$ is almost independent of temperature ($\chi_0 = -3.9(3) \times 10^{-9} \text{ m}^3 \text{ kg}^{-1}$, diamagnetic). It is nearly threefold that of $\alpha\text{-Ge}$ ($-1.328 \times 10^{-9} \text{ m}^3 \text{ kg}^{-1}$), probably as a result of the ‘structural increment’ from molecular ring currents²⁵. The electrical resistivity of cold-pressed powder is quite large ($\rho(300 \text{ K}) = 0.17 \text{ } \Omega \text{ m}$) and increases with decreasing temperature ($1.1 \text{ } \Omega \text{ m}$ at 77 K; $2.2 \text{ } \Omega \text{ m}$ at 4.2 K). However, it is not possible to extract a reliable value for the energy gap from the current resistivity data (Supplementary Figs S2 and S3). The optical absorption spectrum of $\square_{24}\text{Ge}_{136}$, obtained from diffuse reflectance measurements shows a broad band with peaks at 1.60 and 1.30 eV with an absorption edge at about 1.10 eV and a sharp shoulder at lower energies (Supplementary Fig. S4). Subsequent analyses indicated a direct optical bandgap of 0.6 eV, consistent with reported theoretical calculations, in contrast to the indirect bandgap of $\alpha\text{-Ge}$ ($E_g = 0.74 \text{ eV}$).

Differential scanning calorimetry analysis in combination with XRPD shows that, on heating under inert conditions, $\square_{24}\text{Ge}_{136}$ exothermically and irreversibly converts at 420 °C to the cubic $\alpha\text{-Ge}$ (diamond structure), and is therefore a new metastable form of germanium.

The clathrate-II structure of $\square_{24}\text{Ge}_{136}$, discussed in this work in the context of intermetallic phases, was first found in gas hydrates²⁸ and is structurally related to the previously known clathrasile compound dodecasil-3C (ref. 29).

The three-dimensional network consists of tetrahedrally coordinated Ge atoms $\text{Ge}(\text{Ge}_{4/2})_4$, which form hexakaidecahedral [$5^{12}6^4$] Ge_{28} and pentagonal dodecahedral [5^{12}] Ge_{20} cages (Fig. 4a). Each unit cell contains 16 dodecahedra and 8 hexakaidecahedra, which provide space for 24 cage positions in a fully filled clathrate, giving

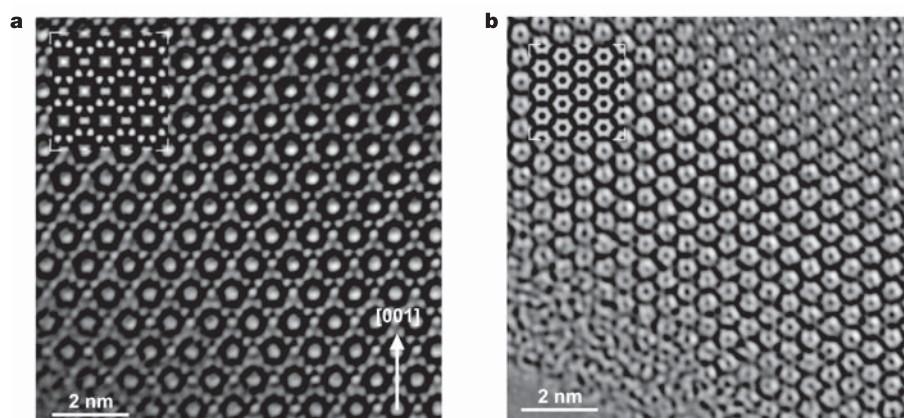


Figure 2 | High-resolution transmission electron microscopy of $\square_{24}\text{Ge}_{136}$. HRTEM micrographs for zone axes [110] (**a**) and [111] (**b**). Simulated images are inserted; the defocus Δf was -50 nm and the specimen thicknesses were $t = 4.3 \text{ nm}$ (**a**) and $t = 5.3 \text{ nm}$ (**b**).

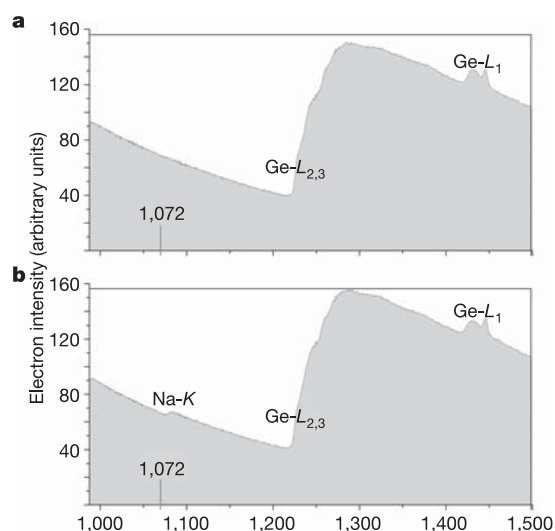


Figure 3 | Chemical analysis of $\square_{24}\text{Ge}_{136}$ by EELS. **a**, EEL spectrum of a $\square_{24}\text{Ge}_{136}$ micro-crystallite (in [100] orientation) showing solely $L_{2,3}$ and L_1 edges of germanium at 1,217 and 1,414 eV, respectively. **b**, Spectrum of the amorphous Ge minority phase showing additionally a weak Na-K edge.

rise to the formula $\square_{24}\text{Ge}_{136}$. Each hexakaidecahedron is surrounded by a super-tetrahedron of 16 dodecahedra whose edges run along the face diagonals of the cubic unit cell (Fig. 4b). The observed Ge–Ge distances in $\square_{24}\text{Ge}_{136}$ are 2.435(1), 2.443(1), 2.445(1) and 2.492(1) Å. The average distance of 2.453 Å closely resembles the Ge–Ge bond distance in α -Ge (2.446 Å) and is significantly smaller than the average Ge–Ge distances for filled $\text{M}_x\text{Ge}_{136}$ clathrates, such as 2.494 Å for $\text{Cs}_8\text{Na}_{16}\text{Ge}_{136}$ (ref. 27).

In addition, the significantly smaller cell volume of $\square_{24}\text{Ge}_{136}$ than in the known filled $\text{M}_x\text{Ge}_{136}$ clathrates is in contrast to $\text{Na}_x\text{Si}_{136}$, in which the unit cell volume does not change appreciably with filling^{8,9}. However, it is consistent with a model of a semiconductor empty clathrate-II in which shorter Ge–Ge distances result from empty antibonding states, in contrast with the filled clathrate-II model in which the electrons from the metal guests supposedly occupy the Ge–Ge antibonding conduction bands. Furthermore, semiconducting $\text{Na}_x\text{Ge}_{136}$ ($x > 0$) should contain additional vacancies in the Ge framework to balance the charge of the alkali atoms in the cages, as discussed in detail for the clathrate-I type³⁰. Such anionic defects should be located at the framework positions of the Ge3 atoms, which form the six-membered rings of the hexakaidecahedra. Although the six-membered rings are not exactly planar in

$\square_{24}\text{Ge}_{136}$ (the sum of the angles is 718.5°), they bear most of the strain in the framework. In fact, the thermal displacement parameters in the isotropic refinement for Ge3 are even slightly smaller than for Ge2 and Ge1, thus negating the possibility of vacancies at this site.

The mild oxidation of Zintl anions resulting in the formation of a previously unknown metastable crystalline elemental phase is demonstrated here. As a result, a new elemental germanium semiconductor with a direct optical bandgap is now available in bulk quantities. Furthermore, the use of quaternary ammonium-based ILs as solvents in the mild oxidation of Zintl anions offers unique opportunities for the preparation of new intermetallic phases with open framework structures analogous to those of silicates and aluminophosphates. Analogous to the chemistry of zeolites and aluminosilicates, new anionic open framework structures can also be expected by varying the synthesis conditions (for example temperature, organic cations, templates and oxidizers, and anions) as well as by aliovalent substitution. The inherent thermal stability of the IL (decomposition temperature about 330 °C) permits access to low and moderate temperatures at which kinetically stabilized phases can be obtained. There is no obvious direct structural relationship between the Ge_9 units and the $\square_{24}\text{Ge}_{136}$ framework structure, so the assembly and oxidation may proceed through intermediate structures that are yet to be determined. Investigations are currently in progress to explain the steps involved in the chemical reactions in the IL. We are currently extending the soft-chemical route to explore whether this is a general route to silicon and germanium structures, as well as using other soluble Zintl phases and anions. So far we have synthesized bulk material of the clathrate-I phases $\text{K}_{8-x}\text{Si}_{46}$ and $\text{K}_8\text{Ge}_{44}\square_2$ and the filled clathrate-II phase $\text{K}_8\text{Ge}_{136}$ under similar conditions. The availability of bulk quantities from the above method, as well as the scalability of a solution method to thin-film fabrication, also provides the possibility of processing framework materials for different applications.

METHODS

Synthesis. A precursor with the nominal composition $\text{NaGe}_{2.25}$ was prepared by melting high-purity elements at 1,050 °C within sealed Ta containers in a tubular furnace for about 1 h, followed by annealing for two days at 800 °C. XRPD patterns of the precursor alloy indicated the formation of Na_4Ge_9 . The ILs were prepared by thoroughly mixing stoichiometric amounts (1:1) of purified dodecyltrimethylammonium chloride and high-purity (sublimed) AlCl_3 in dry glass vessels. According to differential scanning calorimetry (heating rate 10 K min^{−1}) the pure IL melts at 42 °C and does not decompose up to 300 °C. Only a slight colour change from yellow to bright orange was observed after 72 h at 300 °C, whereas the solution turned black at 330 °C, indicating decomposition of the IL. On addition and stirring of the precursor powders (typically 0.1 g $\text{NaGe}_{2.25}$ in 3 g IL) at 50 °C the melt turned to a dark green suspension, indicating the presence of oxidized Ge_5^{3-} anions. The resulting suspension was heated

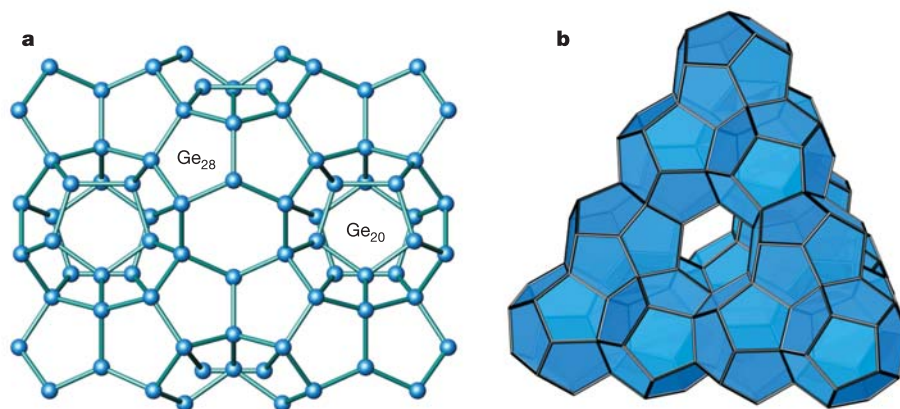


Figure 4 | Structure of empty clathrate-II $\square_{24}\text{Ge}_{136}$. **a**, Unit cell viewed along the [110] direction. The structure contains two basic motifs, dodecahedra with 20 Ge atoms and hexakaidecahedra with 28 Ge atoms. **b**, Sixteen dodecahedra, connected by common faces, build up a super-tetrahedron, whose edges run parallel to the face diagonal of the cubic unit cell.

slowly and kept at 300 °C for 36 h under inert atmosphere in a closed glass container. The presence of water in the IL or oxygen gas in the reaction vessel severely disrupts the reaction. After reaction, dark grey polycrystalline $\square_{24}\text{Ge}_{136}$ was formed and the IL suspension turned clear and colourless. The polycrystalline product was easily separated from the IL by several washings with acetone and water; it was then dried under vacuum. All preparative manipulations and handling of the reactants and the reaction melt were performed under an inert atmosphere in an argon-filled glovebox. Manipulations after the reaction were performed in air. The calculated reaction yield of the air-stable product was about 70%, on the basis of the amount of Na_4Ge_9 used.

Electron microscopy. SAED, HRTEM, energy-dispersive X-ray spectroscopy and EELS were performed with a FEI Tecnai G2-F30 electron microscope ($C_s = 1.2\text{ mm}$) operated at 300 kV, equipped with a Gatan GIF2001 imaging filter and a Gatan UltraScan 1000 charge-coupled device camera. The samples were ground in an agate mortar, dispersed in n-butanol and spread over a holey carbon film. SAED patterns of the clathrate particles in crystal orientations [100], [110], [111], [112] and [221] were recorded, as well as HRTEM defocus-series for zone axes [110] and [111]. HRTEM images shown in Fig. 2 were taken around the Scherzer optimum focus. EEL spectra were measured in different orientations [100], [110] and [111]. Before EELS, the crystalline particles were investigated by SAED and HRTEM and confirmed to be the clathrate phase. Image simulation was performed with the multislice formalism of the EMS program package.

Magnetization and electrical resistivity. The magnetization was determined in a superconducting quantum interference device (SQUID) magnetometer ($2\text{ mT} \leq \mu_0 H \leq 7\text{ T}$; $1.8\text{ K} < T < 400\text{ K}$). The susceptibility was corrected for the sample holder and minor ferromagnetic and paramagnetic impurities. Electrical resistivity (zero field; $3.8\text{ K} < T < 320\text{ K}$) was measured on a cuboid of cold pressed powder by a conventional direct-current four-probe method. Samples were always handled under an argon or helium atmosphere.

X-ray powder diffraction. Measurements of powder samples were performed with a STOE STADI MP diffractometer (Bragg–Brentano geometry, $\text{Cu K}\alpha_1$ radiation). Lattice parameter refinement was performed with LaB_6 as internal standard. All structure refinements were performed with the WinCSD program package.

Diffuse reflectance measurements. Optical diffuse reflectance spectra were collected at room temperature with a Varian CARY-500 Scan UV-VIS-NIR spectrophotometer equipped with a Praying-Mantis diffuse reflectance accessory. A white standard of polytetrafluoroethylene was used as a reference. Absorption spectra were calculated from the reflectance data with the Kubelka–Munk function. The optical bandgap was estimated by applying Tauc plots.

Received 23 March; accepted 8 August 2006.

- Kasper, J. S., Hagemmüller, P., Pouchard, M. & Cros, C. Clathrate structure of silicon $\text{Na}_8\text{Si}_{136}$ and $\text{Na}_x\text{Si}_{136}$ ($x < 11$). *Science* **150**, 1713–1714 (1965).
- Trikalitis, P. N., Rangan, K. K., Bakas, T. & Kanatzidis, M. G. Varied pore organization in mesostructured semiconductors based on the $[\text{SnSe}_4]^{4-}$ anion. *Nature* **410**, 671–675 (2001).
- Cohn, J. L., Nolas, G. S., Fessatidis, V., Metcalf, T. H. & Slack, G. A. Glasslike heat conduction in high-mobility crystalline semiconductors. *Phys. Rev. Lett.* **82**, 779–782 (1999).
- Connétable, D. *et al.* Superconductivity in doped sp^3 semiconductors: The case of the clathrates. *Phys. Rev. Lett.* **91**, 247001 (2003).
- Paschen, S. *et al.* Towards strongly correlated semimetals: $\text{U}_2\text{Ru}_2\text{Sn}$ and $\text{Eu}_8\text{Ga}_{16}\text{Ge}_{30}$. *J. Phys. Chem. Solids* **63**, 1183–1188 (2002).
- Moriguchi, K., Munetoh, S. & Shintani, A. First-principles study of $\text{Si}_{34-x}\text{Ge}_x$ clathrates: Direct wide-gap semiconductors in Si–Ge alloys. *Phys. Rev. B* **62**, 7138–7143 (2000).
- Adams, G. B., O’Keeffe, M., Demkov, A. A., Sankey, O. F. & Huang, Y.-M. Wide-band-gap Si in open fourfold-coordinated clathrate structures. *Phys. Rev. B* **49**, 8048–8053 (1994).
- Gryko, J. *et al.* Low-density framework form of crystalline silicon with a wide optical band gap. *Phys. Rev. B* **62**, 7707–7710 (2000).
- Ammar, A. *et al.* On the clathrate form of elemental silicon, Si_{136} : preparation and characterisation of $\text{Na}_x\text{Si}_{136}$ ($x \rightarrow 0$). *Solid State Sci.* **6**, 393–400 (2004).
- Cundy, C. S. & Cox, P. A. The hydrothermal synthesis of zeolites: History and development from the earliest days to the present time. *Chem. Rev.* **103**, 663–701 (2003).
- Feng, P., Bu, X. & Zheng, N. The interface chemistry between chalcogenide clusters and open framework chalcogenides. *Acc. Chem. Res.* **38**, 293–303 (2005).
- Grüttner, A., Nesper, R. & Von Schnering, H. G. Novel metastable germanium modifications *allo-Ge* and *4H-Ge* from $\text{Li}_{17}\text{Ge}_{12}$. *Angew. Chem. Int. Edn Engl.* **21**, 912–913 (1982).
- Von Schnering, H. G., Schwarz, M. & Nesper, R. The lithium sodium silicide Li_3NaSi_6 and the formation of *allo-silicon*. *J. Less-Common Metals* **137**, 297–310 (1988).
- Wilson, M. & McMillan, P. F. Crystal-liquid phase relations in silicon at negative pressure. *Phys. Rev. Lett.* **90**, 135703 (2003).
- Munetoh, S., Moriguchi, K., Kamei, K., Shintani, A. & Motooka, T. Epitaxial growth of a low-density framework form of crystalline silicon: A molecular-dynamics study. *Phys. Rev. Lett.* **86**, 4879–4882 (2001).
- Dong, J. & Sankey, O. F. Theoretical study of two expanded phases of crystalline germanium: clathrate-I and clathrate-II. *J. Phys. Condens. Matter* **11**, 6129–6145 (1999).
- de A. A. Soler-Illia, G. J., Sanchez, C., Lebeau, B. & Patarin, J. Chemical strategies to design textured materials: from microporous and mesoporous oxides to nanonetworks and hierarchical structures. *Chem. Rev.* **102**, 4093–4138 (2002).
- Von Schnering, H. G. *et al.* The cluster anion Si_9^{4-} . *Angew. Chem. Int. Edn Engl.* **37**, 2359–2361 (1998).
- Quéneau, V. & Sevov, S. C. Ge_9^{4-} : A deltahedral Zintl ion made in the solid state. *Angew. Chem. Int. Edn Engl.* **36**, 1754–1756 (1997).
- Downie, C., Tang, Z. & Guloy, A. M. An unprecedented $_{10}[\text{Ge}_9]^{2-}$ polymer: A link between molecular Zintl clusters and solid-state phases. *Angew. Chem. Int. Edn Engl.* **39**, 338–340 (2000).
- Downie, C., Mao, J.-G., Parmar, H. & Guloy, A. M. The role of sequestering agents in the formation and structure of germanium anion cluster polymers. *Inorg. Chem.* **43**, 1992–1997 (2004).
- McGrath, K. M. Phase behavior of dodecyltrimethylammonium bromide/water mixtures. *Langmuir* **11**, 1835–1839 (1995).
- Cooper, E. R. *et al.* Ionic liquids and eutectic mixtures as solvent and template in synthesis of zeolite analogues. *Nature* **430**, 1012–1016 (2004).
- Grovenstein, E. Jr & Stevenson, R. W. Carbanions. III. Cleavage of tetraalkylammonium halides by sodium in liquid ammonia. *J. Am. Chem. Soc.* **81**, 4850–4857 (1959).
- Cros, C., Pouchard, M. & Hagemmüller, P. Sur une nouvelle famille de clathrates minéraux isotopes des hydrates de gaz et de liquides. Interprétation des résultats obtenus. *J. Solid State Chem.* **2**, 570–581 (1970).
- Gryko, J. *et al.* Electron structure and temperature-dependent shifts in ^{133}Cs NMR spectra of $\text{Cs}_8\text{Ge}_{136}$ clathrate. *Phys. Rev. B* **71**, 115208 (2005).
- Bobev, S. & Sevov, S. C. Synthesis and characterization of stable clathrates of silicon and germanium: $\text{Cs}_8\text{Na}_{16}\text{Si}_{136}$ and $\text{Cs}_8\text{Na}_{16}\text{Ge}_{136}$. *J. Am. Chem. Soc.* **121**, 3795–3796 (1999).
- Von Stackelberg, M. & Müller, H. R. Feste Gashydrate. II. Struktur und Raumchemie. *Z. Elektrochem.* **58**, 25–39 (1954).
- Gies, H., Liebau, F. & Gerke, H. ‘Dodecasile’—eine neue Reihe polytyper Einschlussverbindungen von SiO_2 . *Angew. Chem.* **94**, 214–215 (1982).
- Von Schnering, H. G. Zintl phases: Principles of structure and bonding. *Bol. Soc. Chil. Quim.* **33**, 41–57 (1988).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Kniep and U. Schwarz for comments, R. Cardoso-Gil for XRPD measurements, G. Auffermann and U. Schmidt for chemical analysis, S. Müller for differential scanning calorimetry measurements, R. Koban for resistivity and susceptibility measurements, and P. Simon and W. Carrillo-Cabrera for discussion. A.M.G. thanks the PRF-ACS and the R. A. Welch Foundation for financial support.

Author Contributions All authors contributed equally to this work. A.M.G., Z.T. and M.B. developed the synthetic method and prepared the clathrate material. Yu.G. solved the structure from X-ray diffraction data. R.R. performed electron microscopy investigations. W.S. made measurements of electrical conductivity and magnetic susceptibility, Z.T. performed diffuse reflectance measurements. All authors discussed the results and contributed to the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Yu.G. (grin@cpfs.mpg.de) and A.M.G. (aguloy@uh.edu).

LETTERS

The cause of the fragile relationship between the Pacific El Niño and the Atlantic Niño

Ping Chang¹, Yue Fang¹, R. Saravanan², Link Ji¹ & Howard Seidel¹

El Niño, the most prominent climate fluctuation at seasonal-to-interannual timescales, has long been known to have a remote impact on climate variability in the tropical Atlantic Ocean, but a robust influence is found only in the northern tropical Atlantic region¹. Fluctuations in the equatorial Atlantic are dominated by the Atlantic Niño^{2,3}, a phenomenon analogous to El Niño, characterized by irregular episodes of anomalous warming during the boreal summer. The Atlantic Niño strongly affects seasonal climate prediction in African countries bordering the Gulf of Guinea^{4,5}. The relationship between El Niño and the Atlantic Niño is ambiguous and inconsistent. Here we combine observational and modelling analysis to show that the fragile relationship is a result of destructive interference between atmospheric and oceanic processes in response to El Niño. The net effect of El Niño on the Atlantic Niño depends not only on the atmospheric response that propagates the El Niño signal to the tropical Atlantic, but also on a dynamic ocean–atmosphere interaction in the equatorial Atlantic that works against the atmospheric response. These results emphasize the importance of having an improved ocean-observing system in the tropical Atlantic, because our ability to predict the Atlantic Niño will depend not only on our knowledge of conditions in the tropical Pacific, but also on an accurate estimate of the state of the upper ocean in the equatorial Atlantic.

There is clear observational evidence that the anomalous warming in the eastern equatorial Pacific during an El Niño produces a warming signal in the troposphere as a result of increased atmospheric heating⁶, which then propagates rapidly eastward in the form of an equatorial Kelvin wave and westward in the form of a Rossby wave, as predicted by simple theoretical models^{7,8} (Fig. 1a). The tropospheric warming stabilizes the environment, causing reduced moist convection, which acts in conditionally unstable environments to redistribute energy from the boundary layer to the free troposphere. As a result, boundary layer energy, primarily in the form of latent heat that originates via evaporation from the ocean surface, will accumulate⁹. This then provides a ‘back pressure’ that reduces evaporation from the ocean surface to the boundary layer, leading to a warming of the ocean mixed layer. This so-called ‘tropospheric temperature mechanism’ is arguably an important driving mechanism for the tropical El Niño teleconnection^{10,11}. In the deep tropical Atlantic, it works most effectively from late boreal winter to early spring when the sea surface temperature (SST) is warmest seasonally and the Intertropical Convergence Zone is closest to the Equator. The overall reduction of precipitation in the tropical Atlantic basin observed following El Niños is consistent with the weakening of moist convection¹². Modelling studies^{11,13} consistently show that in the absence of ocean dynamics a basinwide warming of the tropical Atlantic follows an El Niño.

However, the regressed SST response derived from observations

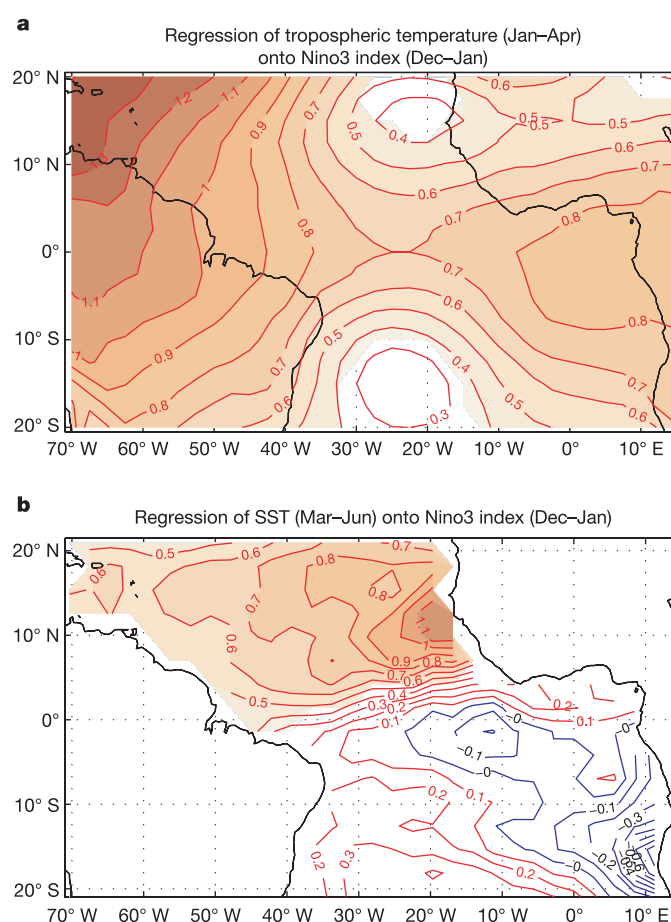


Figure 1 | Tropospheric and surface temperature response of the tropical Atlantic to El Niño revealed by a regression analysis. **a**, The regressed anomalous tropospheric temperature averaged between atmospheric pressure levels of 800 mbar and 200 mbar over January–April against the previous boreal winter, December–January, NINO3 SST (SST anomaly averaged over 150°W–90°W and 5°S–5°N). **b**, The similar regression of the observed March–June SST anomaly. The SST response is expected to lag the tropospheric temperature response because it is driven by surface heat fluxes. The mid-tropospheric temperatures were derived from the 43-year (1958–2000) reanalysis product, ERA40, of the European Centre for Medium-Range Weather Forecasts (ECMWF); <http://www.ecmwf.int/research/era> and the SST is from the Reynolds Optimum Interpolation SST²² during the same period. The coloured shading shows areas that exceed a 99% significance level based on Student's *t*-test.

¹Department of Oceanography, ²Department of Atmospheric Sciences, Texas A&M University, College Station, Texas 77843, USA.

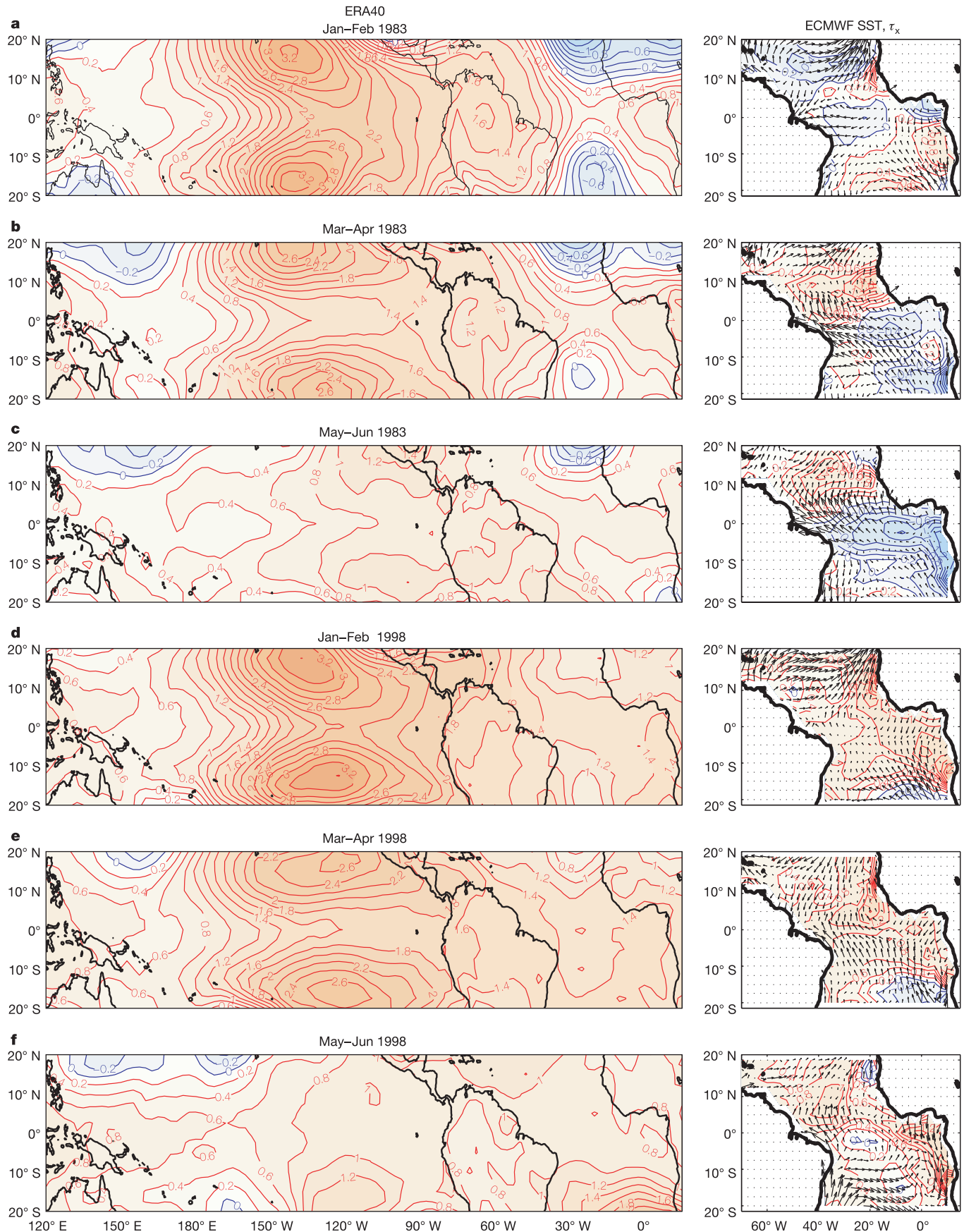


Figure 2 | Atmospheric response to the 1982/83 and 1997/98 El Niños. a–c, 1982/83; d–f, 1997/98. The left panels show the middle troposphere temperature anomaly averaged between atmospheric pressure levels of 800 mbar and 200 mbar, based on the ECMWF ERA40 reanalysis product. The right panels show the observed SST and surface wind stress (τ_x) anomaly from the same ECMWF and Reynolds SST products.

does not show a basinwide warming (Fig. 1b), as one would expect from the tropospheric temperature mechanism. The robust warming is found only in the north tropical Atlantic. This may be attributed partially to the fact that the tropospheric temperature mechanism works more efficiently in this region because of the warmer mean SST (ref. 11) and that El Niños can cause a weakening in the northeasterly trade wind, resulting in a further reduction in evaporative heat loss and thus enhancing the warming¹². In the equatorial and south tropical Atlantic regions, however, the observed response is more perplexing: it tends to be opposite to that in the troposphere, albeit less statistically significant.

The lack of a robust and consistent response of the equatorial SST to El Niño can be made more evident by contrasting the tropospheric and surface response to the 1982/83 and 1997/98 El Niños, two of the strongest events in the instrumental record (Fig. 2). In spite of the well-defined tropospheric warming in both events, the surface responses are drastically different along the Equator and along the southern African coast: persistent surface warming was observed following the 1997/98 El Niño, consistent with the tropospheric temperature mechanism, while a cooling condition prevails in the wake of the 1982/83 El Niño. We wondered what physical factors cause the equatorial Atlantic to respond differently to these El Niños.

First, we considered the surface wind response in the western tropical Atlantic region. In the case of the 1982/83 El Niño, there is a well-defined easterly wind anomaly that persists from the late boreal winter to the early boreal summer in this region (Fig. 2). Such a wind

response is apparently much weaker during the 1997/98 El Niño. The relationship between the wind variability over the western tropical Atlantic and El Niño has previously been noted^{14–16}. A correlation analysis between the observed NINO3 SST time series and the zonal wind stress anomaly reveals an extended area along the western equatorial Atlantic where the correlation coefficient is significantly negative for the months following the peak of El Niño (Fig. 3a). A closer examination reveals that all the El Niño events that did not produce a warming in the equatorial Atlantic were accompanied by anomalously strong easterly winds in this region, while those El Niños that did produce a warming did not have such strong wind anomalies (Fig. 3b).

A previous modelling study¹⁶ suggested that El Niño may lead to cooling in the equatorial Atlantic. Dynamically, the equatorial cooling can be understood in terms of the Bjerknes feedback where, in this case, an easterly wind stress anomaly along the Equator acts to shoal the thermocline, enhancing the stratification and at the same time increasing the vertical entrainment rate on the eastern

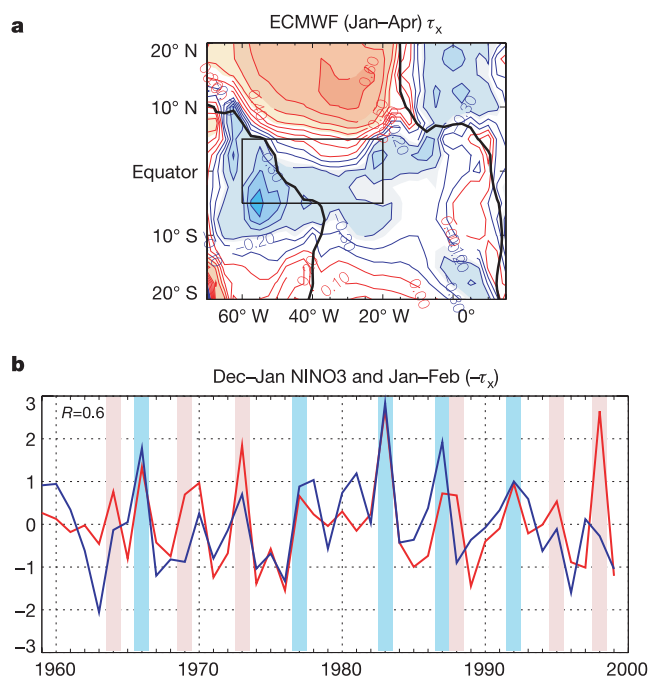


Figure 3 | Observed response of tropical Atlantic zonal wind stress to El Niño. **a**, Correlation between the December–January NINO3 index and the zonal wind stress anomaly from the ECMWF ERA-40 reanalysis product over the tropical Atlantic sector in the following January–April from 1958–2000. The blue contours indicate negative and the red contours positive correlation. The colour shade shows areas that exceed a 95% significance level based on Student's *t*-test. **b**, Normalized time series of the January–February wind stress index (with reversed sign) over the western equatorial region (60°W–20°W and 5°S–5°N) indicated by the rectangle in **a**, and the December–January NINO3 index. The correlation coefficient between the two indices is 0.6. The blue vertical bars indicate those El Niño events (1965/66, 1976/77, 1982/83, 1986/87 and the 1991/92 El Niño) that produced a strong easterly wind anomaly and the red indicate those (1963/64, 1968/69, 1972/73, 1987/88, 1994/95 and the 1997/98 El Niños) that did not.

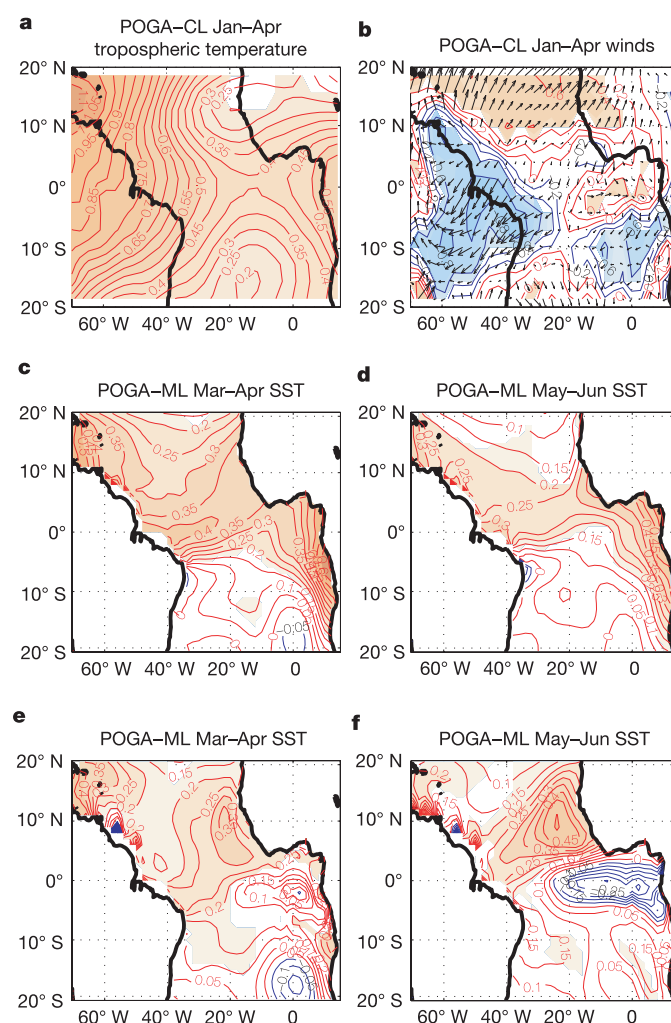


Figure 4 | The simulated tropical Atlantic response to El Niño in three ensembles of numerical experiments. **a**, **b**, Ensemble averaged regressions against the December–January NINO3 index from the POGA-CL runs for the January–April troposphere temperature (°C) averaged between atmospheric pressure levels of 800 mbar and 200 mbar (**a**) and surface winds (vectors; **b**), where the blue contours show the negative correlation of the zonal surface-wind with the NINO3 index and the red contours the positive. (**c**–**f**), Regressions of the ensemble averaged SST in March–April and May–June against the December–January NINO3 index from POGA-ML and POGA-RG runs. The colour shade shows areas that exceed a 99% significance level based on Student's *t*-test.

side of the basin, producing anomalous cooling at the surface. The atmosphere responds to this cooling by further strengthening the wind anomaly, forming a positive feedback between the ocean and atmosphere. We hypothesize that the cooling produced by the Bjerknes feedback competes with the tropospheric-temperature-induced warming. This competition is the main cause of the fragile relationship between the Pacific El Niño and the Atlantic Niño. Borrowing an analogy from wave theory, we refer to this process as 'destructive interference'.

It is difficult to test the destructive interference hypothesis solely on the basis of observational analysis, because observations always capture coupled variability and provide only a single realization of stochastic climate phenomena. We therefore turn to ensembles of coupled- and uncoupled-climate-model experiments (see Methods for model descriptions). The first ensemble of experiments—hereafter referred to as POGA-CL (see Methods)—is designed to examine the direct tropospheric and surface response of the atmosphere to El Niño. Figure 4a and b shows the atmospheric model ensemble mean response. It captures not only the observed structure of the tropospheric warming associated with El Niño (Fig. 1a), but also the surface wind response (Fig. 3a). In particular, the easterly wind anomaly in the western basin that is critically important to the Bjerknes feedback is well-simulated, indicating that the direct atmospheric response to El Niño is already sowing the seeds of the destructive interference.

We then conducted two additional ensembles of coupled-model experiments (hereafter referred to as POGA-ML and POGA-RG, respectively; see Methods). These experiments are specifically designed to test the destructive interference hypothesis by including both the thermodynamic and dynamic ocean-atmosphere interactions in the simulations. POGA-ML experiments, in which only the thermodynamic interaction is permitted, result in a surface warming taking place shortly after the peak of El Niño. The warming persists through the summer months along the Equator and along the southern African coast, as anticipated by the tropospheric temperature mechanism (Fig. 4c, d). Some of the warming along and to the north of the Equator can also be attributed to wind-induced latent heat flux changes¹². In contrast, POGA-RG experiments, in which the dynamic interaction is also permitted, reveal a much weaker boreal spring equatorial warming and a cooling tendency during the early boreal summer (Fig. 4e, f). This cooling effect is attributed to the dynamic ocean response to the easterly surface wind anomaly associated with the warming in the troposphere. Further analyses indicate that the vertical advection of heat is primarily responsible for the cooling. A recent comprehensive coupled-climate study reports a similar finding¹⁷. Therefore, these modelling results support the destructive interference hypothesis as a plausible cause of the fragile relationship between the Pacific El Niño and the Atlantic Niño.

It is unclear what causes the winds in the western equatorial Atlantic to respond strongly to some El Niños, but not others, in spite of a significant overall correlation. One factor may be related to the pre-existing tropical Atlantic SST condition prior to an El Niño¹⁸. Model experiments where the annual cycle of Atlantic SST in the POGA-CL runs was replaced by the observed Atlantic SST show that a pre-existing warm SST anomaly may substantially weaken the wind response, resulting in a noticeable drop in the overall correlation between the western Atlantic wind index and NINO3. Another factor is related to stochasticity of the atmosphere. We noted that the correlation between the wind index derived from an individual ensemble member of the POGA-CL runs and the NINO3 can decrease from the ensemble mean correlation of over 0.8 to about 0.6 for individual members. Additionally, the wind response can depend on the atmospheric heating structure and duration, which can vary considerably from one El Niño to another.

The Atlantic Niño is not a simple passive response to the Pacific El Niño, but a complex one involving destructive interference between

Pacific remote influence and Atlantic ocean-atmosphere feedback, making it a challenge for seasonal climate prediction. The SST anomaly in the equatorial Atlantic—a key climate variable for seasonal climate forecasts—depends on the relative dominance of the tropospheric-temperature-induced heating compared to the dynamic ocean-atmosphere feedback. Therefore, a successful forecast of SST not only requires that climate models simulate both processes accurately, but also requires an accurate estimate of the atmospheric state, both in terms of temperature and surface wind response to El Niño, and of the upper ocean state. The latter points to the necessity of an ocean observing system in the tropical Atlantic because an accurate estimate of the upper ocean state is crucial for determining the strength of the dynamic ocean-atmosphere feedback. In contrast, the ocean dynamics appear less critical in the north tropical Atlantic, for which even an atmospheric general circulation model coupled to a simple slab ocean is useful in forecasting seasonal SST anomalies¹⁹.

METHODS

Models. The atmospheric model is the Community Climate Model version 3 (CCM3) developed at the National Center for Atmospheric Research with the standard configuration of T42 spectral truncation in the horizontal and 18 hybrid levels in the vertical. It incorporates a comprehensive suite of physical parameterizations including a non-local boundary layer parameterization and improved radiative and convection parameterizations²⁰. The coupled model is comprised of the CCM3 as the atmospheric component and an extended 1.5-layer reduced-gravity ocean model that has been used extensively in the study of the Pacific El Niño²¹ and also applied to the study of the Atlantic Niño³. The ocean model has a resolution of 2° in longitude by 1° in latitude and is coupled to the CCM3 in the tropical Atlantic sector (between 30°S and 30°N), outside which the atmospheric model is forced with observed SSTs.

POGA-CL experiments. The Pacific Ocean-Global Atmosphere (POGA)-Climatological (CL) ensemble simulation consists of nine runs where the CCM3 was forced with observed SSTs from 1950 to 1995 in the Pacific and with the annual cycle of SST in the Atlantic. Each ensemble member differs only slightly in its atmospheric initial condition.

POGA-ML experiments. The POGA-Mix Layer (ML) ensemble simulation consists of 12 runs where, in the Pacific sector, the coupled model was forced with observed SSTs from 1980 to 2000. In the tropical Atlantic sector where the model is coupled, the ocean component had all the dynamic processes (except diffusion) disabled, so that the ocean model essentially degenerates to a slab ocean where SST changes are completely determined by atmospheric surface heat fluxes. As in POGA-CL, each ensemble member differs only slightly in its atmospheric initial condition.

POGA-RG experiments. The POGA-Reduced Gravity (RG) ensemble simulation is identical in set-up to POGA-ML, except that all the dynamic processes of the reduced-gravity ocean model are retained.

Received 3 November 2005; accepted 6 July 2006.

- Enfield, D. B. & Mayer, D. A. Tropical Atlantic SST variability and its relation to El Niño–Southern Oscillation. *J. Geophys. Res.* **102**, 929–945 (1997).
- Merle, J. Variabilité thermique annuelle et interannuelle de l'océan Atlantique équatorial Est. L'hypothèse d'un "El Niño" Atlantique. *Oceanol. Acta* **3**, 209–220 (1980).
- Zebiak, S. E. Air–sea interaction in the equatorial Atlantic region. *J. Clim.* **6**, 1567–1586 (1993).
- Goddard, L. & Mason, S. J. Sensitivity of seasonal climate forecasts to persisted SST anomalies. *Clim. Dyn.* **19**, 619–632 (2002).
- Giannini, A., Saravanan, R. & Chang, P. Oceanic forcing of Sahel rainfall on interannual to interdecadal time scales. *Science* **302**, 1027–1030 (2003).
- Yulaeva, E. & Wallace, J. M. Sensitivity of seasonal climate forecasts to persisted SST anomalies. *J. Clim.* **7**, 1719–1736 (1994).
- Gill, A. E. Some simple solutions for heat-induced tropical circulation. *Q. J. R. Meteorol. Soc.* **106**, 447–462 (1980).
- Saravanan, R. & Chang, P. Interaction between tropical Atlantic variability and El Niño–Southern Oscillation. *J. Clim.* **13**, 2177–2194 (2000).
- Brown, R. G. & Bretherton, C. S. A test of the strict quasi-equilibrium theory on long time and space scales. *J. Atmos. Sci.* **54**, 624–638 (1997).
- Chiang, J. C. H. & Sobel, A. H. Tropical tropospheric temperature variations caused by ENSO and their influence on the remote tropical climate. *J. Clim.* **15**, 2616–2631 (2002).
- Chiang, J. C. H. & Lintner, B. R. Mechanisms of remote tropical surface warming during El Niño. *J. Clim.* **18**, 4130–4149 (2005).

12. Giannini, A., Chiang, J. C. H., Cane, M. A., Kushnir, Y. & Seager, R. The ENSO teleconnection to the tropical Atlantic Ocean: contributions of the remote and local SSTs to rainfall variability in the tropical Americas. *J. Clim.* **15**, 4530–4544 (2001).
13. Alexander, M. & Scott, J. The influence of ENSO on air–sea interaction in the Atlantic. *Geophys. Res. Lett.* **29**, doi:10.1029/2001GL014347 (2002).
14. Delecluse, P., Servain, J., Levy, C., Arpe, K. & Bengtsson, L. On the connection between the 1984 Atlantic warm event and the 1982–1983 ENSO. *Tellus A* **46**, 448–464 (1994).
15. Carton, J. A. & Huang, B. Warm events in the tropical Atlantic. *J. Phys. Oceanogr.* **24**, 888–903 (1994).
16. Latif, M. & Barnett, T. P. Interactions of the tropical oceans. *J. Clim.* **8**, 952–964 (1995).
17. Huang, B. Remotely forced variability in the tropical Atlantic Ocean. *Clim. Dyn.* **23**, 133–152 (2004).
18. Giannini, A., Saravanan, R. & Chang, P. The preconditioning of tropical Atlantic variability in the development of the ENSO teleconnection: implications for the prediction of Nordeste rainfall. *Clim. Dyn.* **22**, 839–855 (2004).
19. Chang, P., Saravanan, R. & Ji, L. Tropical Atlantic seasonal predictability: the roles of El Niño remote influence and thermodynamic air–sea feedback. *Geophys. Res. Lett.* **30**, 1501–1504 (2003).
20. Kiehl, J. T., Hack, J. J., Bonan, G. B., Boville, B. A., Williamson, D. L. & Rasch, P. J. The National Center for Atmospheric Research Community Climate Model: CCM3. *J. Clim.* **11**, 1131–1149 (1998).
21. Zebiak, S. E. & Cane, M. A. A model El Niño–Southern Oscillation. *Mon. Weath. Rev.* **115**, 2262–2278 (1987).
22. Reynolds, R. W. & Smith, T. M. Improved global sea surface temperature analysis using optimum interpolation. *J. Clim.* **7**, 929–948 (1994).

Acknowledgements We thank J. C. H. Chiang for a discussion about the tropospheric temperature mechanism. This work was supported by the NOAA Climate and Global Change Program and by the NSF Climate Dynamics Program.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.C. (ping@tamu.edu).

Acceleration of Greenland ice mass loss in spring 2004

Isabella Velicogna^{1,2} & John Wahr¹

In 2001 the Intergovernmental Panel on Climate Change projected the contribution to sea level rise from the Greenland ice sheet to be between -0.02 and $+0.09$ m from 1990 to 2100 (ref. 1). However, recent work^{2–4} has suggested that the ice sheet responds more quickly to climate perturbations than previously thought, particularly near the coast. Here we use a satellite gravity survey by the Gravity Recovery and Climate Experiment (GRACE) conducted from April 2002 to April 2006 to provide an independent estimate of the contribution of Greenland ice mass loss to sea level change. We detect an ice mass loss of $248 \pm 36 \text{ km}^3 \text{ yr}^{-1}$, equivalent to a global sea level rise of $0.5 \pm 0.1 \text{ mm yr}^{-1}$. The rate of ice loss increased by 250 per cent between the periods April 2002 to April 2004 and May 2004 to April 2006, almost entirely due to accelerated rates of ice loss in southern Greenland; the rate of mass loss in north Greenland was almost constant. Continued monitoring will be needed to identify any future changes in the rate of ice loss in Greenland.

The Greenland mass balance in a warming climate is a competition between increased precipitation caused by greater oceanic evaporation, and a combination of increased melting at the ice sheet surface and increased glacial discharge at the coasts. All these trends have been confirmed in recent studies. Regional climate models, supported by *in situ* observations, suggest that both accumulation and melting have increased during the past decade, with melting increasing faster than accumulation⁵. These surface mass balance estimates are consistent with radar altimeter measurements during 1992–2003 that show interior growth^{6,7}, and with laser altimeter observations that show thinning in the 1990s at low elevations⁸ where increased melting is probably more important than increased accumulation. Laser altimeter and satellite radar imaging observations over the last decade have shown accelerated glacial mass loss at the ice sheet margins^{2–4,9,10}.

Here we estimated changes in Greenland mass using an independent technique based on data from the GRACE satellite mission. GRACE, launched in March 2002 by NASA and Deutsches Zentrum für Luft- und Raumfahrt, is mapping the Earth's gravity field every month during its 8–9-year lifetime¹¹. In a previous study a Greenland mass loss of $82 \pm 22 \text{ km}^3 \text{ yr}^{-1}$ was estimated using monthly GRACE gravity solutions for the period April 2002–July 2004¹². Here, we extended that analysis to include solutions to the end of April 2006. GRACE provides a comprehensive survey of the entire ice sheet, free from the issue of incomplete spatial sampling and other limitations that characterize competing techniques. The primary limitations of GRACE are that it cannot provide spatial resolution finer than a few hundred kilometres, and that it is particularly sensitive to post-glacial rebound (PGR, the viscoelastic response of the solid Earth to glacial unloading over the last several thousand years).

Each GRACE monthly gravity solution consists of spherical harmonic (Stokes) coefficients C_{lm} and S_{lm} , to degree l and order m both ≤ 120 . Here, we used CSR Release 1 (<http://podaac-www.jpl.nasa.gov/grace/>) solutions for 42 months between April 2002 and April 2006. The C_{20} coefficients show large variability, so we replaced them with values derived from satellite laser ranging¹³. We used the Stokes coefficients to estimate monthly mass changes of the entire Greenland ice sheet, and of South and North Greenland separately (defined here as the regions south and north of 73.25°). We constructed an averaging function for each region using a method that minimizes the combined measurement error and signal leakage (Fig. 1)¹⁴. GRACE does not recover $l = 1$ Stokes coefficients (representing displacements of the Earth's centre of mass), so we removed those coefficients from the averaging function. We convolved the GRACE Stokes coefficients with these averaging functions to obtain monthly mass change estimates. We simultaneously fitted a trend and annually and semiannually varying terms to each time series.

Before the trends can be interpreted as mass loss estimates, they must be scaled and corrected for contamination from other geophysical signals, and uncertainty estimates must be derived. Procedures for scaling and for estimating uncertainties are described below. For sources of contamination we considered changes in the distribution of (1) continental water storage outside Greenland, (2) water in the ocean, (3) atmospheric mass, and (4) the PGR signal in the solid Earth. Contamination from (3) and (4) comes mostly from mass variability directly above (3) or below (4) the ice sheet. Leakage from (1) and (2) occurs because our averaging functions extended outside Greenland. This latter leakage is increased because our omission of $l = 1$ terms caused the averaging functions to have small-amplitude tails extending around the globe.

To estimate the leakage from (1) and (2), we used global land water storage output from the Global Land Data Assimilation System model¹⁵, and sea floor pressure fields from a baroclinic ocean

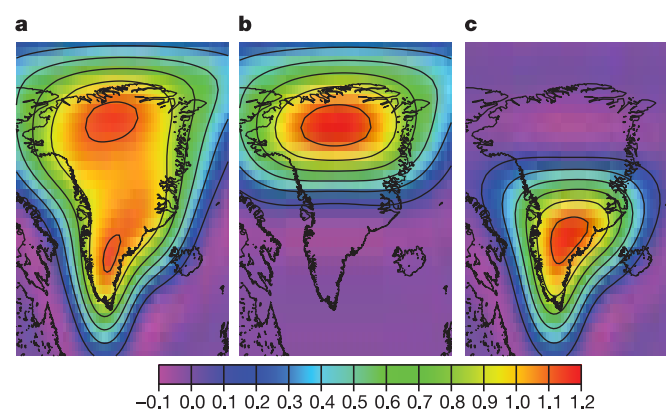


Figure 1 | Averaging functions. Shown are the (unscaled, dimensionless) averaging functions used to estimate the change in total Greenland mass (a), and in the mass of North (b) and South Greenland (c) separately.

¹Department of Physics and CIRES, University of Colorado, Boulder, Colorado 80309-0390, USA. ²Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109-8099, USA.

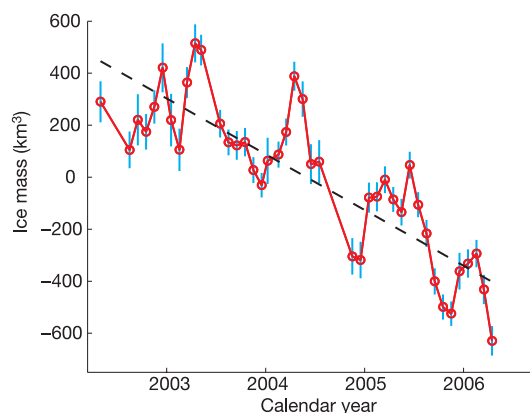


Figure 2 | Greenland GRACE monthly mass solutions. Shown is the GRACE solution for the entire Greenland ice sheet, for April 2002 to April 2006, after scaling the results and removing the mean. The blue error bars include only the contributions from uncertainties in the GRACE gravity fields, and represent 68.3% confidence intervals¹⁸. Also shown is the best-fitting linear trend (dotted line). The results shown here have not been corrected for PGR or for the effects of hydrological or oceanic leakage.

model¹⁶ forced by atmospheric winds and pressure. We added a uniform layer to the global ocean at every time step to conserve the total land + ocean mass. We convolved monthly surface mass estimates from these models with our Greenland averaging functions, and fitted a trend and annually and semiannually varying terms to the results. We used the trend as our estimate of hydrological + oceanographic leakage. Because long-period variability is difficult to model, we adopt a conservative uncertainty estimate equal to $\pm$ the leakage estimate itself.

Figure 2 shows the monthly (scaled) GRACE estimates for all Greenland, before removing the hydrological and oceanographic leakage. There is a clear decrease in mass during this 4-year period. Interpreting the trend as due entirely to a change in ice, and subtracting the leakage trend, we inferred an ice volume decrease of $240 \pm 12 \text{ km}^3 \text{ yr}^{-1}$. The trend obtained without removing the leakage is $224 \pm 8 \text{ km}^3 \text{ yr}^{-1}$. The ± 12 uncertainty is the root sum square (RSS) of the ± 8 gravity field error and the uncertainty in the leakage estimate.

Contamination from the atmosphere and from PGR must be evaluated separately. The GRACE project uses meteorological fields from the European Centre for Medium-Range Weather Forecasts (ECMWF) to remove atmospheric effects from the raw data before constructing gravity fields. We evaluated the probable effects of errors in the meteorological fields by comparing ECMWF pressure fields with pressure observations from stations in the World Meteorological Organization catalogue and from Greenland automatic weather stations¹⁷. The variance of atmospheric errors is less than 3% of the GRACE variance, implying a negligible contribution to the trend uncertainty.

The PGR signal is indistinguishable from a linear trend in ice mass, and must be independently modelled and removed (see below). We obtained a PGR contribution to the GRACE estimate of total Greenland ice decrease of $-8 \pm 21 \text{ km}^3 \text{ yr}^{-1}$. When this PGR contribution was subtracted from the GRACE-minus-leakage trend, we obtained $248 \pm 36 \text{ km}^3 \text{ yr}^{-1}$ as our final estimate of the decrease in total Greenland ice during April 2002–April 2006. The uncertainty is the RSS of the uncertainties in the GRACE fit, in the hydrology + ocean leakage, and in the PGR contribution, with an additional overall 5% error from the uncertainty in the scaling factor. This rate of ice loss corresponds to $0.5 \pm 0.1 \text{ mm yr}^{-1}$ of global sea level rise.

Figure 3 shows the monthly GRACE results after applying the same analysis to North and South Greenland separately. A mass loss is clearly evident in each region, but the South Greenland trend is especially notable. After correcting for the effects of contamination from hydrological, oceanographic, and PGR signals, we obtained ice

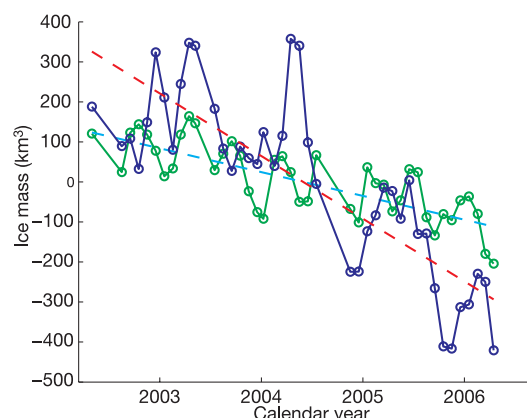


Figure 3 | North and south Greenland GRACE monthly mass solutions.

This is shown as in Fig. 2, but for South Greenland (blue monthly values with the best-fitting trend shown as a red dotted line) and North Greenland (green monthly values with the best-fitting trend shown as a cyan dotted line) separately.

losses of $161 \pm 24 \text{ km}^3 \text{ yr}^{-1}$ and $83 \pm 18 \text{ km}^3 \text{ yr}^{-1}$ for South and North Greenland respectively, during April 2002–April 2006. The mass loss in the south was twice that in the north during these four years. Our results show a significant increase in the rate of Greenland mass loss, starting in spring 2004. A fit to the GRACE results for all Greenland before and after April 2004 yielded ice loss trends of $104 \pm 54 \text{ km}^3 \text{ yr}^{-1}$ during April 2002–April 2004, and $342 \pm 66 \text{ km}^3 \text{ yr}^{-1}$ during May 2004–April 2006.

This corresponds to a 250% increase in the ice loss rate after April 2004. The acceleration occurred mostly in the south. There, the ice loss rate increased from $20 \pm 26 \text{ km}^3 \text{ yr}^{-1}$ during April 2002–April 2004 to $246 \pm 36 \text{ km}^3 \text{ yr}^{-1}$ during May 2004–April 2006. For North Greenland the ice loss rate was about the same during April 2002–April 2004 ($80 \pm 28 \text{ km}^3 \text{ yr}^{-1}$) as during May 2004–April 2006 ($90 \pm 28 \text{ km}^3 \text{ yr}^{-1}$). Changes in mass loss rates are evident in Fig. 4, which shows best-fitting trends for moving two-year data spans, each obtained by simultaneously solving for a trend and annually and semi-annually varying terms. Figure 4 shows that the South Greenland trends increase as the two-year period passes through spring 2004. The North Greenland trends are relatively constant over this time period.

Because only four years of data are available at present, GRACE is not yet able to tell us whether the 250% increase in the Greenland trend is a true long-term increase in ice loss such as might be caused by accelerated glacial discharge, or is an interannual variation in accumulation or melting. However, the timing of the GRACE acceleration is consistent with the dramatic 2004 acceleration of the Kangerdlugssuaq and Helheim glaciers in southeast Greenland that was detected using satellite radar observations⁴. Furthermore, independent estimates of Greenland surface mass balance variability through the end of 2004 (ref. 5) show no evidence of accelerated surface mass loss with the timing and amplitude of the GRACE results. The implication is that the increased mass loss rate inferred from GRACE probably reflects accelerated flow of glacial ice into the ocean from South Greenland, and suggests that this increased flow is probably the dominant mechanism at present in controlling the overall mass balance of the ice sheet. This mass loss component does not figure into traditional surface mass balance estimates, and is difficult to recover with a radar altimeter given the large altimetry footprint.

Our numerical results are consistent with recent remote sensing estimates that indicate the total Greenland mass loss more than doubled between 1996 and 2005, from $90 \text{ km}^3 \text{ yr}^{-1}$ to $220 \text{ km}^3 \text{ yr}^{-1}$ (ref. 2). The GRACE results provide a completely independent measurement that recovers both the glacial discharge and the surface mass balance component in a single observational result, and allow us to identify the timing of this accelerated mass loss and the region

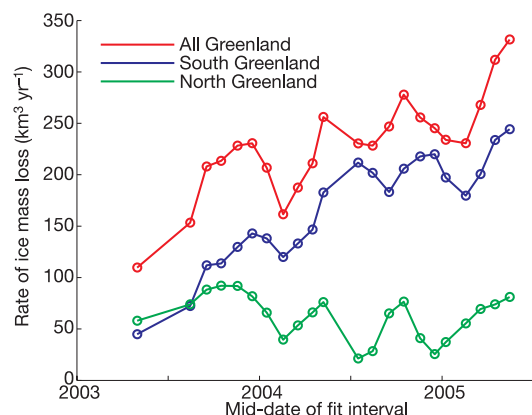


Figure 4 | The best-fitting mass loss trends for each region, as determined for moving two-year data spans. The dates on the x axis indicate the mid-times of each two-year period.

(South Greenland) from which it originated. Mass loss in North Greenland was relatively constant during this time period. There are, however, indications based on radar interferometric surveys that the glacial accelerations already occurring in southern Greenland may be in the process of spreading into regions further north². Continued monitoring with GRACE will help identify any future increase in the rate of ice loss in this region.

METHODS

Scaling. Our optimal averaging functions (Fig. 1) have values less than 1.0 over most of their respective regions, and extend outside those regions. They thus give results that are biased low. To recover unbiased mass estimates for all Greenland we scaled its averaging function so that when applied to a uniform 1-cm ice change over all regions within a few hundred kilometres of the Greenland coast, but zero in the interior, it returns an average Greenland value of 1 cm. We obtained a scaling factor of ~ 2 . This choice of mass distribution is motivated by laser altimeter data suggesting that the largest Greenland mass changes are concentrated at the edges¹⁰. If, instead, we had scaled the averaging function to reproduce the mass from 1 cm of ice spread evenly over all Greenland, the scale factor would be reduced by 5%. This 5% difference is included in our overall uncertainty estimates. To determine scaling factors for South and North Greenland is more complicated because the South Greenland averaging function extends slightly over North Greenland, and vice versa. We applied each averaging function to a uniform mass change over each region individually, and used the four resulting values to determine the linear combination of South and North Greenland results that correctly recovers the mass in each region. Again, these scaling factors have a 5% uncertainty, which we included in our overall uncertainty estimates.

Gravity field uncertainties. We estimated 1σ uncertainties caused by errors in the gravity fields, by convolving our averaging functions with uncertainty estimates for the GRACE Stokes coefficients¹⁸. The uncertainties in each Stokes coefficient were obtained by assuming that the scatter of the 42 monthly values about their best-fitting annual cycle is due entirely to errors, with no contributions from real geophysical signals. This is certainly not true and led to overestimated errors. The convolution with an averaging function assumes that the errors in different Stokes coefficients are uncorrelated. This assumption is not true either, and led to error estimates for the mass results that oversimplify their spatial pattern. The scatter of the non-annually varying mass estimates shows larger amplitudes at high northern latitudes than were expected from our error estimates^{18,19}. This could be due either to non-annually varying geophysical signals, or to error correlations between Stokes coefficients. Comparable results derived from hydrological and oceanographic models show enough similarity with GRACE to suggest that the increased GRACE scatter at high latitudes probably reflects contributions from real geophysical signals, rather than increased errors there¹⁸.

Post-glacial rebound. The two main sources of PGR model error are the ice history and the Earth's viscosity profile. We used two Greenland ice history models: ICE-5G²⁰ and GREEN1²¹. For ice loading outside Greenland we use both ICE-5G and ICE-3G²². We convolved these ice histories with viscoelastic Green's functions for an incompressible Earth¹², constructed using a wide range of plausible two-layer viscosity profiles. We computed a set of Stokes coefficient trends for each Green's function and each ice model, and convolved those trends with the GRACE Greenland averaging functions. We obtained a range of possible PGR contributions to the GRACE mass trends. For our preferred PGR contribution we

used results based on ICE-5G and the VM2 viscosity profile adopted for the construction of ICE-5G (ref. 20). We used the range of values determined from all other viscosity profiles and ice models as our PGR uncertainty estimate.

We concluded that the PGR signal causes an apparent Greenland ice decrease of $-8 \pm 21 \text{ km}^3 \text{ yr}^{-1}$. About half this uncertainty comes from not knowing the viscosity profile, and half comes from the uncertainty in the ice model. The relatively small preferred value ($-8 \text{ km}^3 \text{ yr}^{-1}$) may seem incompatible with the expectation that the Earth beneath Greenland should be rebounding upward at a significant rate owing to the Holocene removal of Greenland ice. But Greenland lies outside the forebulge of the Pleistocene ice sheet in northern Canada, and so there is subsidence caused by the removal of that ice sheet. The $-8 \text{ km}^3 \text{ yr}^{-1}$ value comes from the near-cancellation of $\sim 25 \text{ km}^3 \text{ yr}^{-1}$ signals caused by the Greenland and non-Greenland ice histories. The degree of this cancellation is different in North and South Greenland, where the preferred values are $-9 \text{ km}^3 \text{ yr}^{-1}$ and $+5 \text{ km}^3 \text{ yr}^{-1}$, respectively.

Received 21 June; accepted 14 August 2006.

- Church, J. A. *et al.* in *Climate Change 2001: The Scientific Basis* (eds Houghton, J. T. *et al.*) 639–693 (Cambridge Univ. Press, Cambridge, UK, 2001).
- Rignot, E. & Kanagaratnam, P. Changes in the velocity structure of the Greenland ice sheet. *Science* **311**, 986–990 (2006).
- Howat, I., Joughin, I., Tulaczyk, S. & Gogineni, S. Rapid retreat and acceleration of Helheim Glacier, east Greenland. *Geophys. Res. Lett.* **32**, L22502, doi:10.1029/2005GL024737 (2005).
- Luckman, A., Murray, T., de Lange, R. & Hanna, E. Rapid and synchronous ice-dynamic changes in East Greenland. *Geophys. Res. Lett.* **33**, L03503, doi:10.1029/2005GL025428 (2006).
- Box, J. *et al.* Greenland ice sheet surface mass balance variability (1988–2004) from calibrated Polar MM5 output. *J. Clim.* **19**, 2783–2800 (2006).
- Johannessen, O. *et al.* Recent ice-sheet growth in the interior of Greenland. *Science* **310**, 5750, doi: 10.1126/science.111535 (2005).
- Zwally, J. *et al.* Mass changes of the Greenland and Antarctic ice sheets and shelves and contributions to sea-level rise: 1992–2002. *J. Glaciol.* **51**, 509–527 (2005).
- Krabill, W. *et al.* Greenland ice sheet: High-elevation balance and peripheral thinning. *Science* **289**, doi:10.1126/science.289.5478.428 (2000).
- Abdalati, W. *et al.* Outlet glacier and margin elevation changes: Near-coastal thinning of the Greenland ice sheet. *J. Geophys. Res.* **106**, doi:10.1029/2001JD900192 (2001).
- Krabill, W. *et al.* Greenland Ice Sheet: Increased coastal thinning. *Geophys. Res. Lett.* **31**, L24402, doi:10.1029/2004GL021533 (2004).
- Tapley, B., Bettadpur, S., Watkins, M. & Reigber, C. The Gravity Recovery and Climate Experiment: Mission overview and early results. *Geophys. Res. Lett.* **31**, L09607, doi:10.1029/2004GL019920 (2004).
- Velicogna, I. & Wahr, J. Greenland mass balance from GRACE. *Geophys. Res. Lett.* **32**, L18505, doi:10.1029/2005GL023955 (2005).
- Cheng, M. & Tapley, B. Variations in the Earth's oblateness during the past 28 years. *J. Geophys. Res.* **109**, B09402, doi:10.1029/2004JB003028 (2004).
- Swenson, S., Wahr, J. & Milly, P. C. D. Estimated accuracies of regional water storage anomalies inferred from GRACE. *Wat. Resour. Res.* **39**, 1223, doi:10.1029/2002WR001808 (2003).
- Rodell, M. *et al.* The Global Land Data Assimilation System. *Bull. Am. Meteorol. Soc.* **85**, 381–394 (2004).
- Dobslaw, H. & Thomas, M. Atmospheric induced oceanic tides from ECMWF forecasts. *Geophys. Res. Lett.* **32**, L10615, doi:10.1029/2005GL022990 (2005).
- Steffen, K. *Greenland Climate Network (GC-Net)* (<http://cires.colorado.edu/science/groups/steffen/gcnet>) (2005).
- Wahr, J., Swenson, S. & Velicogna, I. The accuracy of GRACE mass estimates. *Geophys. Res. Lett.* **33**, L06401, doi:10.1029/2004GL019779 (2006).
- Horwath, M. & Dietrich, R. Errors of regional mass variations inferred from GRACE monthly solutions. *Geophys. Res. Lett.* **33**, L07502, doi:10.1029/2005GL025550 (2006).
- Peltier, W. R. Global glacial isostasy and the surface of the ice-age earth: the ICE-5G(VM2) model and GRACE. *Annu. Rev. Earth Planet. Sci.* **32**, 111–149 (2004).
- Fleming, I. & Lambeck, K. Constraints on the Greenland Ice Sheet since the Last Glacial Maximum from sea-level observations and glacial-rebound models. *Quat. Sci. Rev.* **23**, 1053–1077 (2004).
- Tushingham, A. M. & Peltier, W. R. ICE-3G: A new global model of late Pleistocene deglaciation based upon geophysical predictions of postglacial relative sea level change. *J. Geophys. Res.* **96**, 4497–4523 (1991).

Acknowledgements We thank S. Bettadpur, J. Cheng, J. Reis, E. Rignot and M. Watkins for data and advice. This work was supported by NASA's Cryospheric and Solid Earth Programs and by the NSF Office of Polar Programs. This research was carried out in part at the Jet Propulsion Laboratory.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to I.V. (isabella@colorado.edu).

LETTERS

Geological and palaeontological context of a Pliocene juvenile hominin at Dikika, Ethiopia

Jonathan G. Wynn^{1†}, Zeresenay Alemseged², René Bobe³, Denis Geraads⁴, Denné Reed⁵ & Diana C. Roman⁶

Since 1999, the Dikika Research Project (DRP; initiated by Z.A.) has conducted surveys and excavations in badlands that expose Pliocene and Pleistocene sediments south of the Awash River in Ethiopia, between surrounding hominin localities at Hadar¹, Gona² and the Middle Awash region³. Here we report our geological mapping and stratigraphic measurement of the DRP area, and the context of a remarkably well-preserved skeleton of the earliest known juvenile hominin at the Dikika DIK-1 locality⁴. Our mapping of the DRP area permits a complete definition of the hominin-bearing Hadar Formation and provides a cohesive structural and tectonic framework defining its relationships to adjacent strata. Our findings reveal the basin-scale tectonic, depositional and palaeoenvironmental history of the area, as well as a clear taphonomic and palaeontological context for the juvenile hominin. Such data are crucial for understanding the environmental context of human evolution^{5,6}, and can be integrated into larger-scale tectonic and palaeoenvironmental studies^{7,8}. Our basin-scale approach to palaeoenvironments provides a means to elucidate the complex geological history occurring at the scale of temporally and geographically controlled fossil point localities^{3,9–11}, which occur within the rich tectonic and depositional history of the Awash Valley.

The DRP area lies southwest of the Tendaho-Goba'ad discontinuity, which separates two tectonic regimes: the southeast-spreading Ethiopian Rift system (ERS) and the northeast-spreading Red Sea Rift system (RSRS)^{7,8}. Because the DRP area exposes the entirety of the Hadar Formation (>3.8–2.9 million years (Myr) old) and Busidima Formation (2.7–<0.6 Myr old), our observations of local structure and stratigraphy allow reconstruction of the basin-scale tectonic and depositional history (Fig. 1). We find evidence of the DRP area being influenced by both the RSRS and ERS tectonic regimes during migration of the triple junction northward since 11 Myr ago (ref. 8). During deposition of the Hadar Formation, rapid sedimentation (30–90 cm kyr^{−1}) filled the then-active Hadar Basin. Thickening of the Hadar Formation with increasing representation of lacustrine facies occurs northeastwards across a series of northwest-trending syndepositional faults, indicating that local extension of this basin configuration occurred parallel to RSRS extension at ~N 45° E. Our mapping also provides a definitive lower boundary to the Hadar Formation in contact with the uppermost weathered surface of basalt flows, as originally suspected by a previous study¹². Although the previous map showed these as the Afar stratoid series basalt (ASSB), we interpret the flows below the Hadar Formation as the Dahla series basalt (DSB). Unlike the DSB, local flows of the ASSB are syndepositional with the Hadar Formation (often showing features of cooling in the Hadar Lake), and do not extend southwest of a series of northwest-trending faults that cut across the DRP (one flow of the

ASSB is the Kadada Moumou basalt¹; another caps Andalee Ridge). The upper surfaces of two DSB-flows visible below the Hadar Formation each have a deeply weathered residual palaeosol, indicating an extended period of non-deposition before formation of the Hadar Basin (Fig. 2, section G). In the southeast DRP, the uppermost palaeosol is overlain by poorly sorted juvenile volcanic-arenitic sands, which interfinger with lacustrine clays of the lower Hadar Formation further basinward (northeast). The presence of the Ikini Tuff (a correlate of the Wargolo Tuff from the Turkana Basin¹³) indicates that the Hadar Formation sediments extend at least to the age of this tuff (3.8 Myr old; ref. 13) and overlie the weathered surface of the DSB.

The Hadar Formation upper boundary is defined by an angular unconformity to the overlying Busidima Formation² (Fig. 1b; the Busidima unconformity surface, 2.9–2.7 Myr old; ref. 2). Deposition of the Busidima Formation clearly postdates most of the fault offsets. Thus, later than 2.9 Myr ago, deposition in the Hadar Basin ceased and the Hadar Formation strata were uplifted from a depositional position by normal faults parallel to currently active basin-bounding faults of the Yangudi Basin in the Southern Afar Rift (Fig. 1). This period of uplift and erosion marks a similar change in extension direction in the Main Ethiopian Rift from 135° to 105°, indicated by the onset of similarly oriented faulting in the Adama Basin⁸ (Fig. 1). Following 2.7 Myr ago, local sedimentation began in the Busidima Half-graben (filled with sediments of the Busidima Formation), which thickens westwards towards the As Duma border fault. This tectonic setting contrasts strongly with that of the Hadar Basin, and is similar to a series of more recent marginal grabens that extend into the northern Afar along the edge of the western escarpment⁷.

Initiation of sediment accumulation in the Hadar Basin is marked by a sharp transition from deep weathering and erosion of the basaltic basement to rapid deposition of lacustrine clays of the Basal Member that thicken northeast from zero to >60 m. Lake shoreline facies, which are represented by two gastropod-bearing bioclastic sandstones (B-g, Basal gastropodite) near the top of the Basal Member, transition to a subaerial lower-delta-plain and delta-channel facies-association that contains the juvenile hominin⁴ in the lower Sidi Hakoma Member (DIK-1 locality; Fig. 1). This depositional setting, combined with the remarkable preservation of many articulated faunal remains lacking evidence of preburial weathering, most probably indicates rapid deposition during major flood events, burying many fossils as intact corpses (including the juvenile hominin). We have mapped regional depositional environments during emplacement of the Sidi Hakoma Tuff, from shallow lacustrine in the northeast DRP area, to offshore lacustrine delta at Arbosh, subaqueous delta channel at Gango Akidora, and a mappable upper delta plain fluvial channel in upper Arbosh (Fig. 2). The river-dominated

¹School of Geography and Geosciences, University of St Andrews, St Andrews KY16 9AL, Scotland, UK. ²Department of Human Evolution, Max-Planck-Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany. ³Department of Anthropology, University of Georgia, Athens, Georgia 30602-1619, USA. ⁴CNRS UPR 2147, 44 rue de l'Amiral Mouchez, F-75014 Paris, France. ⁵Department of Anthropology, University of Texas at Austin, Austin, Texas 78712, USA. ⁶Department of Geology, University of South Florida, Tampa, Florida 33620, USA. [†]Present address: Department of Geology, University of South Florida, Tampa, Florida 33620, USA (J.G.W.).

delta system in the DIK-1 area transitions to a stable shoreline and lacustrine deposition in the upper Sidi Hakoma through the lower Denen Dora members. The upper Denen Dora through most of the Kada Hadar members shows a return to predominantly fluvial facies, with only brief sedimentological evidence of local transgression of the Hadar Lake, such as that represented by the confetti clay.

The diverse fauna from the excavation site at DIK-1 (ref. 4) were collected from two trough cross-bedded tabular sandstones with basal pebble or grit lags, representing distributary delta-channel facies (Fig. 2). Primates are very rare: in addition to the hominid *Australopithecus afarensis*, only isolated teeth of other primates—*Theropithecus darti* and Cercopithecidae (indeterminate)—were recovered during screening. Carnivores are uncommon except a relatively frequent large otter, *Enhydriodon*. A well-preserved canid cranium may be conspecific with a *Nyctereutes*-like mandible, but is probably not of this genus. A cranium of *Elephas* is more derived than that of *E. ekorensis*, and resembles *E. recki brumpti*. Other molars have similar morphology but fewer plates and are more brachyodont, reminiscent of *E. ekorensis*. Coexistence of these closely related species supports recent views¹⁴ that their evolution is more complex than a single anagenetic lineage. The DIK-1 *Hipparion* is relatively large, but a complete metatarsal is slightly smaller than specimens from Hadar¹⁵. Lower teeth of *Hipparion* often have a large ectostylid, showing that this feature is not restricted to later Pliocene forms. A hexaprotodont hippo with a wide symphysis has second incisors, similar in size or slightly smaller than the third incisors, that fill a gap between teeth size of the two named species in the Hadar Formation¹⁶. Despite within-member size variation among suids, clear metric trends can be demonstrated in members of the genera

Kolpochoerus and *Notochoerus*. *K. afarensis* from the Hadar Formation is larger than *K. deheinzeli* from Aramis, Sagantole and Chad, but specimens from Amado (northeast of Hadar) bridge the size gap between these species. Within the DRP, the size of third molars (and premolars) in *K. afarensis* increases in the Basal Member and the lower Sidi Hakoma Member. The size of third molars of *Notochoerus* does not increase, but a mandible from the DIK-1 locality confirms a previous suggestion that premolars from the base of the Sidi Hakoma Member are still *Nyanzachoerus*-like, whereas those discovered at Denen Dora and Kada Hadar members are significantly reduced¹⁷. The most common bovid at DIK-1 is an impala, slightly smaller and more primitive according to the weaker lamination of its horn-cores than the *Aepyceros shungurae* from the Shungura Formation¹⁸, but more closely related to an unnamed species there¹⁹. Horn-cores from the Sidi Hakoma Member are slightly larger but less compressed than those from the Denen Dora Member. They are much shorter and less spiralled than those of the current form, but also than those from Lothagam²⁰, indicating co-occurrence of two lineages of impalas in the East African Pliocene. A bovid cranium has horn-cores similar to those of *Tragelaphus nakuae* from the Turkana Basin. A similar form from the Denen Dora Member at Hadar was reported¹⁸, but it is also present in the Sidi Hakoma Member. We provisionally refer a single almost straight horn-core to the poorly defined genus *Praedamalis*. It might belong either to the type-species, *P. deturi*, from Laetoli and the Denen Dora Member at Hadar¹⁸ (also found in the Sidi Hakoma Member), or to *P. howelli* from Maka²¹. An alcelaphine, *Damalops sidihakomai*, is represented by a horn-core and teeth of a derived pattern, although its p4 is primitive according to its open lingual valley, as at Maka. Lower molars attributed to *Golunda* from DIK-1

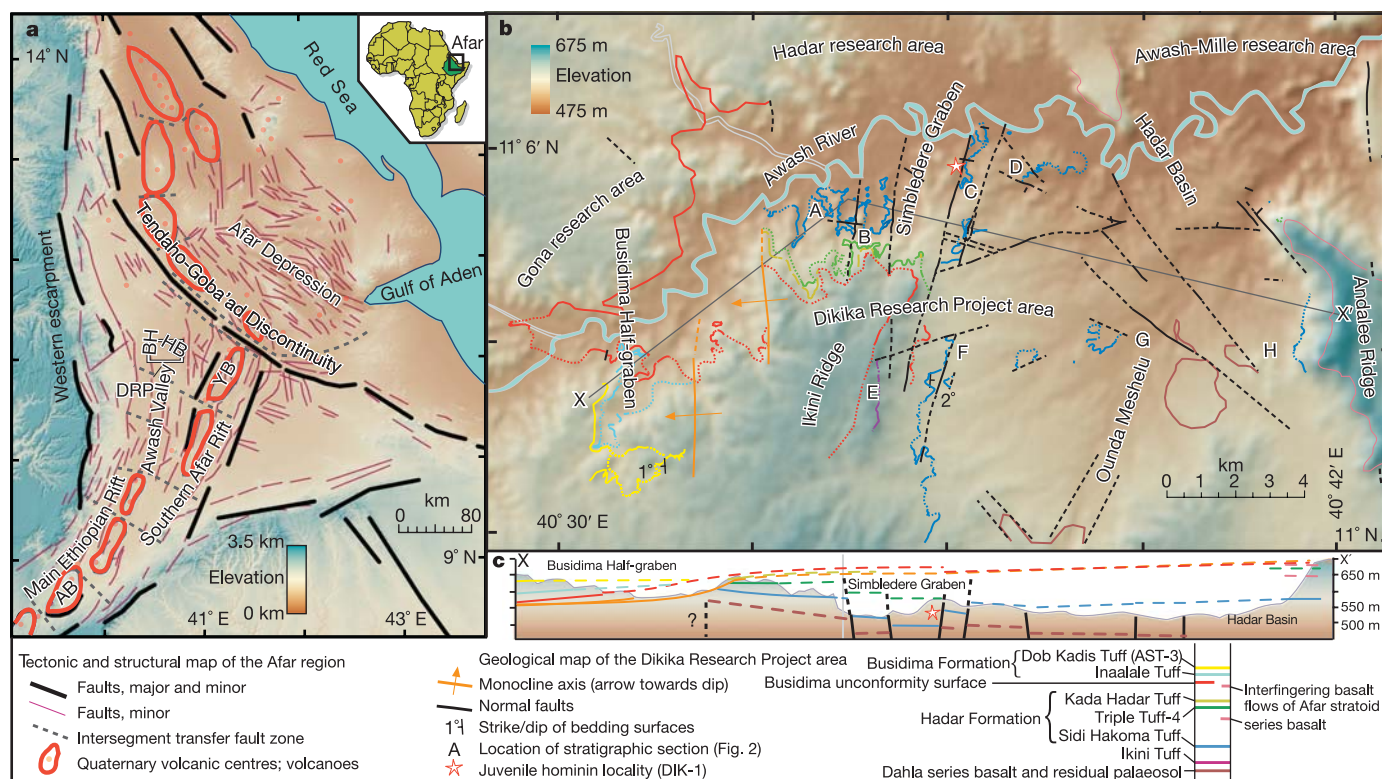


Figure 1 | Maps of the Afar region and of Dikika Research Project area. **a**, Tectonic framework of the Afar region. **b**, Geological map of the eastern DRP area. **c**, Cross-section of the DRP area (approximately the white box in **a**). Inactive basins (both shown normal to extension direction) include Hadar Basin (HB) and Busidima Half-graben (BH), and undescribed basins of the Middle Awash Valley. A cross section along the front of badland exposures in the Awash Valley (line, X–X') shows structural relationships

between the Hadar Formation (deposited in the Pliocene Hadar Basin; thickening northeastward) and the Busidima Formation (deposited in the Busidima Half-graben), providing a simple structural explanation for the Hadar Formation exposures in the Hadar/Dikika area and for the Busidima Formation exposures in the Gona/Asbole area. YB, Yangudi Basin; AB, Adama Basin; AST-3, Artefact Site Tuff at Gona.

closely resemble *G. gurai*, but the one upper molar series does not show the cusp orientation and elongation that characterizes the extant *G. ellioti* and to a lesser extent the fossil *G. gurai*²². This specimen is tentatively referred to a similar genus *Pelomys*, also known from Olduvai and Omo.

The $^{13}\text{C}/^{12}\text{C}$ ratios ($\delta^{13}\text{C} = [({}^{13}\text{C}/{}^{12}\text{C})_{\text{sample}}/({}^{13}\text{C}/{}^{12}\text{C})_{\text{standard}} - 1] \times 1,000$) of carbon in pedogenic carbonate ($\delta^{13}\text{C}_{\text{pc}}$; Fig. 2) from Andedo and Gango Akidora (~3.5–3.1 Myr old) indicate a local increase of C_4 plants after ~3.4 Myr ago, a trend that continues through the Busidima Formation at Gona²³. The sharp contrast between Hadar Formation and Busidima Formation stable isotope values is not unexpected, given the extreme differences in their tectonic and depositional settings described above. A mean $\delta^{13}\text{C}_{\text{pc}}$ of -9.5‰ from a dark-coloured and clayey, poorly drained

palaeosol below sands yielding the juvenile hominin indicates an insignificant contribution from C_4 grasses in a delta plain setting. C_4 grasses, which outcompete C_3 plants in well-drained and seasonally watered aridland soils, increase to nearly one-third of the biomass in the soils developed on delta channel deposits (mean $\delta^{13}\text{C}_{\text{pc}} \approx -6\text{‰}$), and further through the Denen Dora and Kada Hadar members. The abundance of freshwater gastropods, fishes (mostly catfish), hippopotamids, crocodiles and giant tortoises (Table 1) associated with the hominin corroborates the interpretation of a mesic deltaic environment, with nearby permanent water. The DIK-1 fauna includes many species from mesic bush- and tree-associated genera such as *Ugandax*, *Golunda*, and *Millardia*, and possibly *Praomys* and *Pelomys*. This, combined with the absence of open- and xeric-adapted gerbils lends support to wooded

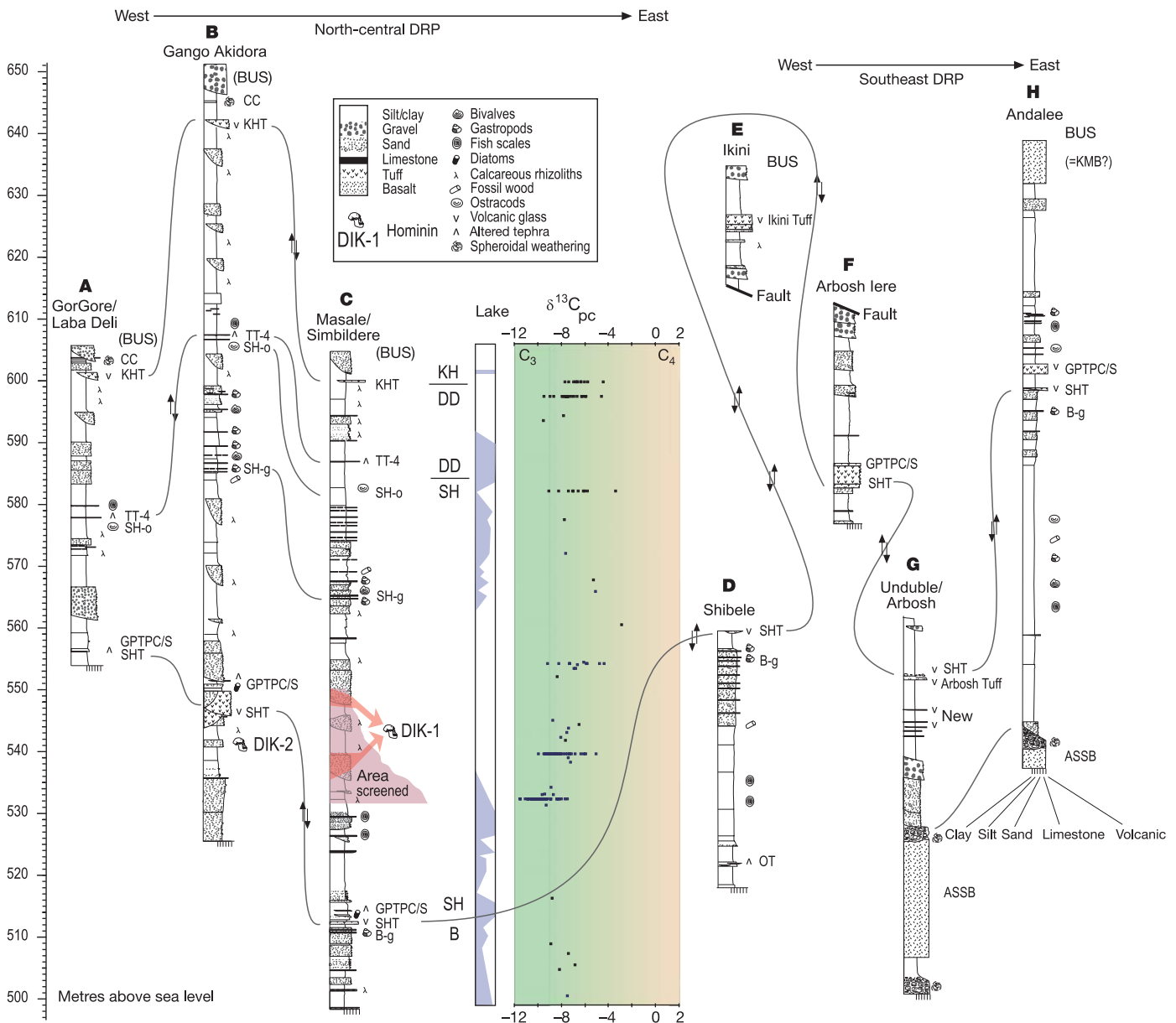


Figure 2 | Stratigraphy of the Hadar Formation in the Dikika Research Project area. Location of sections are shown in Fig. 1. $\delta^{13}\text{C}$ of pedogenic carbonate and lacustrine facies are shown from Gango Akidora, Masele, and Simbildere with respect to section C height. In the north-central DRP (sections A–D) the lower Hadar Formation thickens towards the Hadar Basin centre across syndepositional faults (primarily oriented northwest),

and is offset by predominantly postdepositional faults forming the Simbildere Graben (north–northeast-trending). In southeastern DRP (sections E–H), the Hadar Formation also thickens northeast, particularly the Basal Member, which is deposited on the local sedimentary basement (weathered surface of Dahla series basalt). See Methods for definitions of the stratigraphic markers.

environments. Impalas are the most common antelopes at DIK-1 and in equivalent levels (SH1 sand, Sidi Hakoma Member) at Hadar, a fact that is consistent with the prevalence of C₃ vegetation. However, the high abundance of Alcelaphini (wildebeest relatives) and of the rhinocerotid genus *Ceratotherium*, in addition to the large number of elephants, indicates that open grasslands were an important part of the palaeolandscape. The vertebrate fauna associated with the juvenile hominin therefore suggests a mosaic of mesic habitats. A greater relative abundance of grazing mammals at DIK-1 indicates palaeoenvironments that were more open than those associated with *Ardipithecus* (5.8–4.4 Myr ago; Adu-Asa and Sagantole Formations, Middle Awash region, Ethiopia)^{10,24}, *Australopithecus anamensis* (4.2–4.1 Myr ago; Kanapoi Formation, Turkana Basin, Kenya and Sagantole Formation, Middle Awash region, Ethiopia)^{9,25} and *Kenyanthropus platyops* (3.5 Myr ago; Nachukui Formation, Turkana Basin, Kenya)²⁶. Both the DIK-1 fauna and that from the equivalent interval, SH1, at Hadar indicate mosaic environments. However, the greater proportion of grazing bovids at DIK-1 shows that Dikika had more open habitats than Hadar at that time, probably owing to its depositional position further upstream from the lake. As the fossil vertebrate surveys at Dikika focused on narrow stratigraphic intervals in the lower Sidi Hakoma Member, further work is needed to determine the extent of time-averaging represented by the documented fauna.

Table 1 | Vertebrate fauna of the DIK-1 hominin locality and surroundings

Pisces
<i>Clarias</i> sp.* (24)
Reptilia
<i>Elapidae</i> cf. <i>Naja</i> sp.* (1)
<i>Crocodylus</i> cf. <i>niloticus</i> * (13)
<i>Centrochelys</i> sp.* (24)
Mammalia
Rodentia * (14 total)
<i>Golunda</i> cf. <i>gurai</i> * (6)
<i>Acomys coppensi</i> * (4)
<i>Millardia taiebi</i> * (1)
<i>Pelomys</i> sp.* (1)
<i>Saidomys</i> cf. <i>afarensis</i> * (1)
Primates
<i>Australopithecus afarensis</i> * (1)
<i>Theropithecus darti</i> * (5)
<i>Cercopithecidae</i> indet.* (3)
Carnivora
<i>Herpestidae</i> indet.* (1)
<i>Canidae</i> indet. aff. <i>Nyctereutes</i> sp.*† (2)
<i>Enhydriodon</i> sp. (1)
Proboscidea
<i>Elephas</i> cf. <i>recki brumpti</i> *† (several fragments)
<i>Elephas</i> cf. <i>ekorensis</i> (several fragments)
Perissodactyla
<i>Ceratotherium mauritanicum</i> * (16)
' <i>Hipparion</i> ' cf. <i>hasumense</i> * (10)
Artiodactyla
<i>Hexaprotodon afarensis</i> * (91)
<i>Kolpochoerus afarensis</i> * (8)
<i>Notochoerus euilus</i> * (4)
<i>Nyanzachoerus kanamensis</i> * (3)
<i>Giraffa</i> cf. <i>jumae</i> * (6)
<i>Sivatherium maurusium</i> (3)
<i>Tragelaphus</i> aff. <i>nakuae</i> * (3)
<i>Ugandax</i> sp.* (5)
<i>Praedamalis</i> cf. <i>howelli</i> *† (2)
<i>Damalops</i> sp. (10)
<i>Aepyceros</i> sp.* (17)
<i>Gazella</i> sp.* (17)

The list includes fauna from the immediate vicinity of the DIK-1 site between ~510 and 560 m in section C of Fig. 2; this roughly corresponds to the SH1 submember of the Hadar Formation biostratigraphic terminology. Numbers in parentheses indicate the number of specimens identified in a systematic surface survey and by screening at the site by R.B. and D.G. Numbers in parentheses for Rodentia indicate number of specimens identified in systematic screening for micromammals by D.R. cf., confer (similar to); aff., affinis (related to); indet., indeterminate (not determined).

*Taxa most closely associated with the hominin skeleton of DIK-1.

†New occurrence for the Hadar Formation.

METHODS

Field data in Fig. 1 are superimposed on Shuttle Radar Tomography Mission (SRTM) digital elevation data. The geological map of the DRP outlines stratigraphic marker horizons that delimit the Hadar and Busidima Formations and member boundaries of the Hadar Formation. Major structural features of the ERS include transform-fault zones between sub-basins (Fig. 1; tectonics and structure are detailed according to refs. 7, 8). Active rift segments include the Main Ethiopian Rift, Southern Afar Rift and Afar Depression. Depositional sub-basins with active volcanic centres include the Yangudi Basin, and Adama Basin. Local lithostratigraphy (shown in Fig. 2) is based on a definition of the Hadar Formation²⁷ using which exposures across the Awash River from Dikika were originally mapped²⁷. The Busidima Formation is defined from the Gona area². Hadar Formation member names are: B (Basal; base of sediments–Sidi Hakoma Tuff (SHT)), SH (Sidi Hakoma; SHT–Triple Tuff-4), DD (Denen Dora; TT-4–Kada Hadar Tuff (KHT)), and KH (Kada Hadar; KHT–Busidima unconformity surface (BUS)). Stratigraphic framework is based on tephrostratigraphic correlations that allow direct comparisons between the DRP fauna and others of the Awash Valley (see Supplementary Table 1), where considerable effort has been focused on radiometric dating of volcanic strata interfingering with hominin-bearing sediments^{3,11}. Altered tephra indicated in Fig. 2 are not analysed due to complete weathering of glass shards. Local to regional stratigraphic markers and the method of determining age are: the surface of uppermost DSB flow (>3.8 Myr old), Ikini Tuff (equivalent to Wargolo; 3.8 Myr old; ref. 13), Ounda Leita Tuff (OT; 3.49 Myr old; stratigraphic scaling and sedimentation rate extrapolated from the SH member), Arbosh Tuff (equivalent to Maka Sand Tuff; 3.40 Myr old; ref. 3), B-g (Basal gastropodite; 3.40 Myr old; sedimentation rate extrapolated from the SH member), SHT (equivalent to Tulu Bor Tuff and B-β Tuff; 3.40 Myr old; ref. 28), GPTPC/S (Green Pumiceous Tuff pebble conglomerate/sand; 3.40 Myr old; stratigraphic scaling), Kada Damoumou Basalt (KMB; 3.28 Myr old; ref. 30), SH-g (Sidi Hakoma gastropodite; 3.28 Myr old; stratigraphic scaling), SH-o (Sidi Hakoma ostracodite; 3.23 Myr old; stratigraphic scaling), TT-4 (3.22 ± 0.01 Myr old; ref. 29), KHT (3.18 Myr old; ref. 29), confetti clay (CC; 3.16 Myr ago, stratigraphic scaling) and BUS (previously referred to as the Major Disconformity Surface at Hadar, but named BUS at Gona; beginning 2.9 Myr ago, but in parentheses in Fig. 2 where there may be post-2.7 Myr erosion; ref. 30). Section elevation is determined from the intersection of a prominent tuff (SHT in most sections) and the digital elevation surface from Fig. 1.

Received 24 April; accepted 6 July 2006.

1. Taieb, M., Johanson, D. C., Coppens, Y. & Aronson, J. L. Geological and palaeontological background of Hadar hominid site, Afar, Ethiopia. *Nature* **260**, 289–293 (1976).
2. Quade, J. et al. Paleoenvironments of the earliest stone toolmakers, Gona, Ethiopia. *Geol. Soc. Am. Bull.* **116**, 1529–1544 (2004).
3. White, T. D. et al. New discoveries of *Australopithecus* at Maka in Ethiopia. *Nature* **366**, 261–265 (1993).
4. Alemseged, Z. et al. A juvenile early hominin skeleton from Dikika, Ethiopia. *Nature* doi:10.1038/nature05047 (this issue).
5. Behrensmeyer, A. K. Climate change and human evolution. *Science* **311**, 476–478 (2006).
6. Feibel, C. S. Debating the environmental factors in hominid evolution. *GSA Today* **7**, 1–7 (1997).
7. Beyene, A. & Abdelsalam, M. Tectonics of the Afar Depression: a review and synthesis. *J. Afr. Earth Sci.* **41**, 41–59 (2005).
8. Wolfenden, E., Ebinger, C., Yirgu, G., Deino, A. & Ayelew, D. Evolution of the northern Main Ethiopian rift: birth of a triple junction. *Earth Planet. Sci. Lett.* **224**, 213–228 (2004).
9. White, T. et al. Asa Issie, Aramis and the origin of *Australopithecus*. *Nature* **440**, 883–889 (2006).
10. WoldeGabriel, G., White, T. D., Suwa, G., Renne, P. & de Heinzelin, J. Ecological and temporal placement of early Pliocene hominids at Aramis, Ethiopia. *Nature* **371**, 330–333 (1994).
11. WoldeGabriel, G. et al. Geology and palaeontology of the Late Miocene Middle Awash valley, Afar rift, Ethiopia. *Nature* **412**, 175–178 (2001).
12. Kalb, J. E., Oswald, E. B., Mebrate, A., Tebedge, S. & Jolly, C. J. Stratigraphy of the Awash Group, Middle Awash Valley, Afar, Ethiopia. *Newsl. Stratigr.* **11**, 95–127 (1982).
13. Haileab, B. & Brown, F. H. Turkana Basin–Middle Awash Valley correlations and the age of the Sagantole and Hadar Formations. *J. Hum. Evol.* **22**, 453–468 (1992).
14. Todd, N. E. in *The World of Elephants: Proc. 1st Int. Cong.* (eds Caveratta, G., Gioia, P., Mussi, M. & Palombo, M. R.) 693–697 (Consiglio Nazionale delle Ricerche, Roma, 2001).
15. Bernor, R. L. & Armour-Chelu, M. in *Neogene Paleontology of the Manonga Valley, Tanzania* (ed. Harrison, T.) 219–264 (Plenum, New York, 1997).
16. Boissier, J.-R. & White, T. D. new species of Pliocene Hippopotamidae from the Middle Awash, Ethiopia. *J. Vertebr. Paleontol.* **24**, 474–480 (2004).

17. Cooke, H. B. S. Pliocene–Pleistocene Suidae from Hadar, Ethiopia. *Kirtlandia* **29**, 1–63 (1978).
18. Gentry, A. W. Notes on Bovidae (Mammalia) from the Hadar Formation and from Amado and Geraru, Ethiopia. *Kirtlandia* **33**, 1–30 (1981).
19. Gentry, A. W. in *Les faunes plio-pléistocènes de la basse vallée de l'Omo (Ethiopie), I: Périssodactyles-Artiodactyles (Bovidae)* (eds Coppens, Y. & Howell, F. C.) 119–191 (CNRS, Paris, 1985).
20. Harris, J. M. in *Lothagam: the Dawn of Humanity in Eastern Africa* (eds Leakey, M. G. & Harris, J. M.) 531–579 (Columbia, New York, 2003).
21. Vrba, E. S. New fossils of Alcelaphini and Caprinae (Bovidae: Mammalia) from Awash, Ethiopia, and phylogenetic analysis of Alcelaphini. *Paleont. Afr.* **34**, 127–198 (1997).
22. Sabatier, M. Les rongeurs du site Pliocène à Hominidés de Hadar (Ethiopie). *Palaeovertebrata* **12**, 1–56 (1982).
23. Levin, N., Quade, J., Simpson, S. W., Semaw, S. & Rogers, M. Isotopic evidence for Plio-Pleistocene environmental change at Gona, Ethiopia. *Earth Planet. Sci. Lett.* **219**, 93–110 (2004).
24. Haile-Selassie, Y. et al. Mio-Pliocene mammals from the Middle Awash, Ethiopia. *Geobios.* **37**, 536–552 (2004).
25. Leakey, M. G., Feibel, C. S., McDougall, I. & Walker, A. New four-million-year-old hominid species from Kanapoi and Allia Bay, Kenya. *Nature* **376**, 565–571 (1995).
26. Leakey, M. G. et al. New hominin genus from eastern Africa shows diverse middle Pliocene lineages. *Nature* **410**, 433–440 (2001).
27. Taieb, M., Coppens, Y., Johanson, D. C. & Kalb, J. Dépôts sédimentaires et faunes du Plio-Pléistocène de la basse vallée de l'Awash (Afar central, Ethiopie). *C.R. Acad. Sci. D* **275**, 819–882 (1972).
28. Brown, F. H. & Cerling, T. E. Stratigraphical significance of the Tulu Bor Tuff of the Koobi Fora Formation. *Nature* **299**, 212–215 (1982).
29. Walter, R. C. Age of Lucy and the first family: Laser $^{40}\text{Ar}/^{39}\text{Ar}$ dating of the Denen Dora Member of the Hadar Formation. *Geology* **22**, 6–10 (1994).
30. Kimbel, W. H. et al. Late Pliocene *Homo* and Oldowan stone tools from the Hadar Formation (Kada Hadar Member), Ethiopia. *J. Hum. Evol.* **31**, 549–561 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Haile-Mariam of the Authority for Research and Conservation of Cultural Heritage, and the Afar Regional State for research permission; and G. Assefa and H. Habtemichael for aid in the field. This work would not be possible without the assistance of the many Issa field crew, as well as logistical support from M. Mekonnen, H. Defar, A. Zerihun, A. Takele, A. Kiros. We also thank J.-C. Rage for snake identification, and K. Reed and G. Eck for access to the Hadar catalogue and collections. This research was funded by the National Geographic Society with logistical support from the Institute for Human Origins and the Ministère des Affaires Étrangères. Stable isotope analyses were performed at RSES, Australian National University, Canberra and at FEEA Stable Isotope Laboratory, St. Andrews. Tephrostratigraphic chemical analyses were performed at the University of Oregon Microprobe Laboratory, Eugene, with funds from the Baldwin Memorial Fund.

Author Contributions J.G.W. was the project geologist. D.C.R. did tephrostratigraphy. Fauna were analysed by Z.A. (head of project and palaeoanthropology), R.B. (palaeoenvironments), D.G. (biostratigraphy) and D.R. (micromammals).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.G.W. (jwynn@cas.usf.edu).

A common progenitor for haematopoietic and endothelial lineages in the zebrafish gastrula

Kevin M. Vogel^{1,2,3*}, Suk-Won Jin^{1,3*†}, Gail R. Martin^{2,3} & Didier Y. R. Stainier^{1,3}

It has been proposed that haematopoietic and endothelial cells share a common progenitor, termed the haemangioblast. This idea was initially conceived as a result of the observation that these two cell types develop in close proximity to each other within the embryo^{1,2}. Support for this hypothesis was provided by studies on single-cell-derived colonies that can produce both haematopoietic and endothelial cells *in vitro*^{3–7}. Although these data point towards the existence of a common progenitor for these two lineages, the presence of a bipotential progenitor cell has yet to be demonstrated *in vivo*. Through the construction of single-cell-resolution fate maps of the zebrafish late blastula and gastrula, we demonstrate that individual cells can give rise to both haematopoietic and endothelial cells. These bipotential progenitors arise along the entire extent of the ventral mesoderm and contribute solely to haematopoietic and endothelial cells. We also find that only a subset of haematopoietic and endothelial cells arise from haemangioblasts. The endothelial descendants of the haemangioblasts all clustered in a specific region of the axial vessels regardless of the location of their progenitors. Our results provide *in vivo* evidence supporting the existence of the haemangioblast and reveal distinct features of this cell population.

To investigate whether haematopoietic and endothelial cells arise from a common progenitor, individual cells within the ventral portion of gastrula-stage zebrafish embryos were labelled via laser activation of caged fluorescein dextran. Briefly, single-cell-stage embryos were injected with 2,3-dimethyl 2,3-dinitrobutane (DMNB)-caged fluorescein (FITC) dextran for ubiquitous distribution, and then single cells were activated with an ultraviolet laser at shield stage, 6 h post fertilization (hpf). We selected this time, which is 1 h after the onset of gastrulation movements, because previous fate mapping data indicated that cells are restricted to a single lineage by shield stage^{8,9}. The position of the FITC-labelled cell was immediately digitally recorded from lateral and animal pole views (Fig. 1a). At 30 hpf (Fig. 1b), embryos were scored for descendants of the single labelled cell (FITC; arrows, Fig. 1c, g) that expressed either EGFP from a *flk1* promoter (*flk1*:EGFP; endothelium, Fig. 1d, h) or DsRed from a *gata1* promoter (*gata1*:DsRed; erythrocyte, Fig. 1e, i). Merged images are shown in Fig. 1f, j. Consistent with earlier fate mapping studies^{8,9}, we found that a majority (59.1%) of the labelled cells in the ventral portion of the embryo gave rise to erythrocytes alone (Fig. 2a, b), whereas 20.5% gave rise to endothelial cells alone (Fig. 2a, c). A substantial number of individual cells (12.5%)

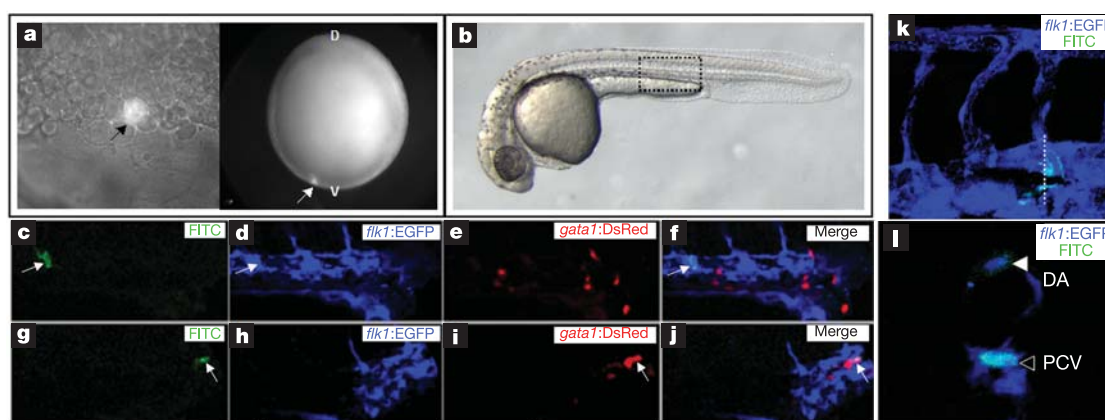


Figure 1 | Single cells at shield stage give rise to only blood and endothelial cells. *Tg(flk1:EGFP)⁸⁴³*; *Tg(gata1:DsRed)* embryos were injected with caged fluorescein dextran at the one-cell stage, and a single cell activated at shield stage (6 hpf). Epifluorescence pictures show the location of a single labelled cell. **a**, Lateral view to determine cell diameters from the margin (left panel). The right panel is an animal pole view showing degrees from dorsal as determined by the distance from the shield (D, dorsal; V, ventral). Embryos were allowed to develop until 30 hpf and then scored for co-localization of activated FITC with EGFP or DsRed. **b**, Brightfield picture of a 30 hpf embryo. The box demarcates the location of endothelial descendants of haemangioblasts. **c–f** and **g–j**, Two separate lateral confocal

sections of a 30 hpf embryo within the region defined by the box in **b**. Embryos were imaged using a $\times 40$ objective with an optical section thickness of $0.58 \mu\text{m}$. **c, g**, Activated FITC showing the descendants of a single, labelled cell. **d, h**, *flk1*:EGFP-labelled endothelial cells. **e, i**, *gata1*:DsRed-labelled erythrocytes. **f, j**, Merge of **c–e**, arrow points to a FITC-labelled endothelial cell. **j**, Merge of **g–i**, showing a different cell in the same embryo as that shown in **c–f**, but in a more superficial focal plane; arrow points to a FITC-labelled erythrocyte. **k**, Lateral confocal view of the location of the endothelial descendants of a haemangioblast. **l**, Transverse vibratome section of **k** showing labelled cells within the dorsal aorta (DA; filled arrowhead) and the posterior cardinal vein (PCV; open arrowhead).

¹Department of Biochemistry and Biophysics, ²Department of Anatomy, and ³Cardiovascular Research Institute, University of California, San Francisco, 1550 Fourth Street, San Francisco, California 94158-2324, USA. [†]Present address: Department of Cell and Molecular Physiology and the Carolina Cardiovascular Biology Center, University of North Carolina at Chapel Hill, 103 Mason Farm Road, Chapel Hill, North Carolina 27599, USA.

*These authors contributed equally to this work.

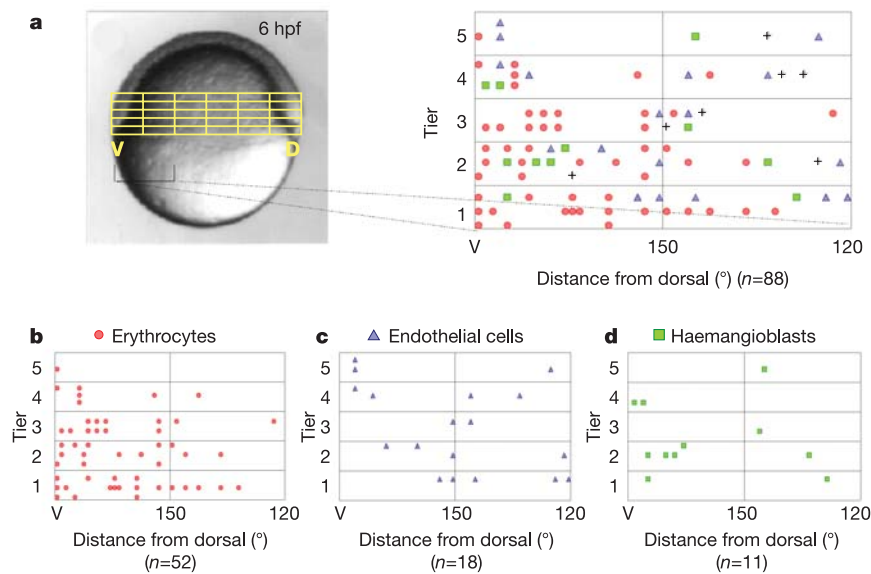


Figure 2 | Single-cell-resolution fate map at shield stage. **a**, Fate map of all cells ($n = 88$) labelled at shield stage (6 hpf) with each shape representing a single labelled cell. Circles represent the cells for which the descendants were only *gata1*:DsRed⁺ erythrocytes (52 cells); triangles represent the cells for which the descendants were only *flk1*:EGFP⁺ endothelial cells (18 cells); squares represent the haemangioblasts (that is, cells for which the

descendants were both *gata1*:DsRed⁺ and *flk1*:EGFP⁺; 11 cells); and crosses represent cells for which the descendants contributed to somites or pectoral fin, but were not erythrocytes or endothelial cells (7 cells). **b**, Location of all FITC-labelled progenitors of *gata1*:DsRed⁺ cells. **c**, Location of all FITC-labelled progenitors of *flk1*:EGFP⁺ cells. **d**, Location of all FITC-labelled haemangioblasts.

specifically gave rise to both lineages (Fig. 2a, d), in accordance with the definition of the haemangioblast. These bipotential haemangioblast cells gave rise to only blood and endothelial cells and not to other lineages, were localized all around the ventral margin of the embryo, and were interspersed with cells fated to become blood or endothelial cells. Previous attempts to identify haemangioblasts in the mouse embryo suggest that they are extremely rare^{7,10}. However, we found that in the early zebrafish gastrula, a substantial percentage of the cells in the ventral margin were haemangioblasts. It is likely that this difference reflects the fact that we focused our fate mapping

experiments on the part of the early gastrula previously determined to give rise to blood and endothelial cells^{8,9}. The anterior–posterior position of endothelial cells has been shown to correspond very broadly to the initial dorsal–ventral position of their progenitors at late blastula stages¹¹. Progenitors of the head vasculature tend to localize dorsally, whereas progenitors of the trunk and tail vasculature tend to localize ventrally. During our lineage studies, we observed a similar correspondence for labelled cells that gave rise to only endothelial cells; however, we found that the endothelial descendants of haemangioblasts did not follow this

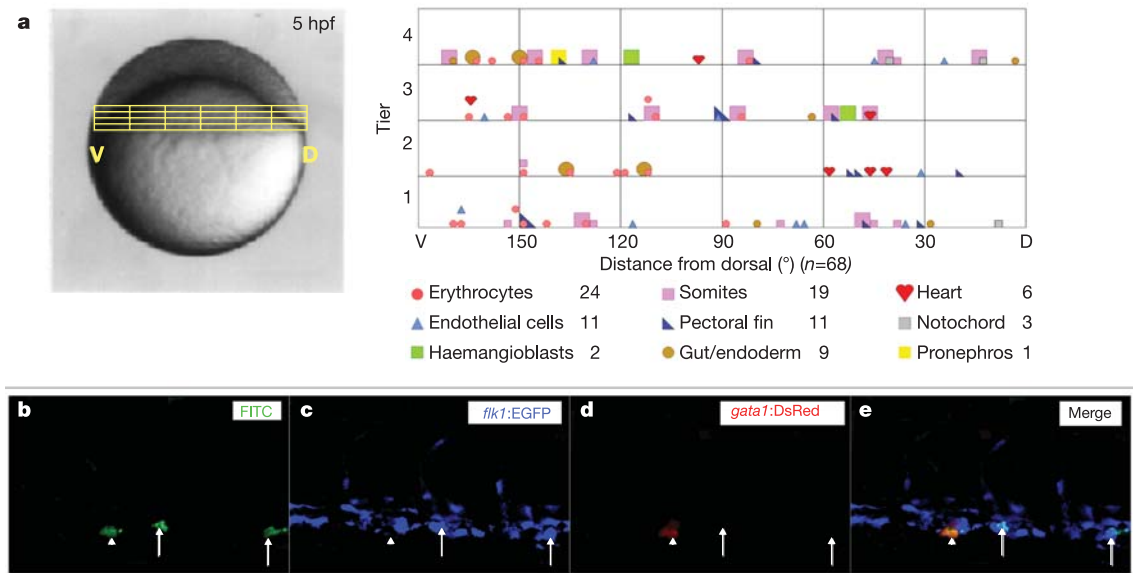


Figure 3 | Single-cell-resolution fate map at 40% epiboly. **a**, Fate map of all cells ($n = 68$) labelled at 40% epiboly (5 hpf) with each shape representing a single labelled cell. Positions on the map at which there are two different shapes (one superimposed on the other) correspond to the locations of single cells for which the descendants were found in two lineages. **b–e**, Single, lateral confocal section of the axial vessels of a 30 hpf

embryo showing the descendants of a single cell labelled at 40% epiboly. The embryo was imaged with a $\times 40$ objective with an optical section thickness of $1.9 \mu\text{m}$. **b**, Descendants of a single FITC-labelled cell. **c**, *flk1*:EGFP-labelled endothelial cells. **d**, *gata1*:DsRed-labelled erythrocytes. **e**, Merge of all channels showing co-localization of activated FITC with *flk1*:EGFP (arrows) and *gata1*:DsRed (arrowhead).

trend. In order to define better the embryonic location of the endothelial descendants of the haemangioblasts, we performed confocal microscopy on 30 hpf embryos in which individual haemangioblasts had been labelled. Notably, the endothelial descendants of the haemangioblasts were consistently located in a specific area dorsal to the yolk extension (box, Fig. 1b), within the axial vessels (Fig. 1f, k), regardless of the initial dorsal–ventral position of their progenitors at shield stage. To examine further their location, we analysed transverse sections at 30 hpf, and observed that they contributed to both the dorsal aorta and posterior cardinal vein (Fig. 1l). The migration of the endothelial descendants of the haemangioblasts to a specific location in the axial vessels, regardless of the dorsal–ventral origin of their progenitors at gastrula stage, might reflect their unique properties.

In order to test further whether all endothelial and blood cells arise from haemangioblasts, and also to investigate the presence and prevalence of haemangioblasts at earlier developmental stages, we performed fate mapping experiments at 40% epiboly (5 hpf). At this stage, one cannot distinguish dorsal from ventral, leading us to survey the entire margin of the embryo. Consistent with previous fate mapping data^{8,9}, cells at this stage were not entirely restricted to a single lineage. For example, cells giving rise to blood or endothelial cells were often found to give rise to other cell types, sometimes apparently even to endodermal cells (Fig. 3a). Haemangioblasts, or their progenitors (that is, cells giving rise to blood, endothelium and a third cell type), were found only rarely (2.9%), suggesting that they might be localized in parts of the embryo we were not able to survey for technical reasons. Notably, our fate mapping data at this earlier stage clearly show that not all endothelial and blood cells arise from a common progenitor (Fig. 3a).

To our knowledge, our data provide the first *in vivo* evidence for the existence of the haemangioblast. At shield stage, these bipotential cells are interspersed along the extent of the ventral mesoderm with cells that give rise to only blood or only endothelial cells. The endothelial descendants of the haemangioblasts were subsequently found in a stereotyped position, dorsal to the yolk extension within the axial vessels of the embryo. On the basis of gene expression analyses¹², this region is thought to be the initial location of definitive haematopoiesis, suggesting that the endothelial descendants of the haemangioblasts might contribute to this process. Thus, the haemangioblasts seem to represent a distinct, specialized cell type in the developing zebrafish embryo.

METHODS

Zebrafish husbandry. Zebrafish (*Danio rerio*) embryos were obtained from a mixed wild-type strain and raised at 28 °C as previously described¹³. Hemizygous transgenic *Tg(flk1:EGFP)^{s843}* (ref. 14) and double hemizygous *Tg(flk1:EGFP)^{s843}; Tg(gata1:DsRed)* (ref. 15) embryos were used.

Immunohistochemistry. Immunohistochemistry was performed as previously described¹⁶. Anti-FITC goat IgG (Molecular Probes) was used at a concentration of 1:100, anti-GFP mouse IgG2a (Molecular Probes) was used at a concentration of 1:500, and anti-DsRed rabbit IgG (BD Biosciences) was used at a concentration of 1:200. Processed samples were mounted in Vectashield (Vector Laboratories) and imaged using a Zeiss LSM5 Pascal confocal microscope.

Lineage tracing. Our lineage tracing strategy was adapted from those previously described^{17,18}. Briefly, double hemizygous *Tg(flk1:EGFP)^{s843}; Tg(gata1:DsRed)* fish were intercrossed and the resulting embryos injected with 4.6 nl of 0.2% DMNB-caged, biotinylated, lysine-fixable fluorescein dextran (10 kDa; Molecular Probes) in 0.2 M KCl, and allowed to develop until 40% epiboly (5 hpf) or shield stage (6 hpf). Fluorescein was activated in a single cell in each embryo using pulses from a 365-nm laser (Photonic Instruments) focused through a ×40 objective on a Zeiss Axioskop 2 microscope. The laser was calibrated in the z-plane and centred in the field of vision by puncturing test holes (approximately 1 µm in diameter) in a foil-coated slide. Additionally, the intensity of the laser was regulated through the use of a neutral density filter. The focus and

intensity of the laser were adjusted such that it was able to puncture only a small point in the foil, and only in the plane of focus. After laser activation of a single cell, each embryo was assessed for supernumerary labelled cells, in the same or different focal planes. Embryos in which more than one cell had been labelled were discarded. The position of the cell was then recorded digitally using a Spot camera. In the case of cells activated at 40% epiboly, the embryos were allowed to develop until shield stage so that their dorsal–ventral position could be determined. Embryos at 30 hpf were fixed overnight in 4% PFA and then permeabilized in 100% methanol. In order to detect FITC, EGFP and DsRed, embryos were incubated in anti-FITC goat immunoglobulin, anti-GFP mouse IgG2a, and anti-DsRed rabbit immunoglobulin, followed by anti-goat Alexa Fluor 568, anti-mouse Alexa Fluor 488, and anti-rabbit Alexa Fluor 648 (Molecular Probes).

Received 17 March; accepted 5 July 2006.

1. Sabin, F. Preliminary note on the differentiation of angioblasts and the method by which they produce blood-vessels, blood-plasma and red blood-cells as seen in the living chick. *Anat. Rec.* **13**, 199–204 (1917).
2. Murray, P. D. F. The development *in vitro* of the blood of the early chick embryo. *Proc. R. Soc. Lond. B* **11**, 497–521 (1932).
3. Choi, K., Kennedy, M., Kazarov, A., Papadimitriou, J. C. & Keller, G. A common precursor for hematopoietic and endothelial cells. *Development* **125**, 725–732 (1998).
4. Schuh, A. C., Faloony, P., Hu, Q. L., Bhimani, M. & Choi, K. *In vitro* hematopoietic and endothelial potential of *flk-1*^{−/−} embryonic stem cells and embryos. *Proc. Natl Acad. Sci. USA* **96**, 2159–2164 (1999).
5. Fehling, H. J. et al. Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation. *Development* **130**, 4217–4227 (2003).
6. Huber, T. L., Kouskoff, V., Fehling, H. J., Palis, J. & Keller, G. Haemangioblast commitment is initiated in the primitive streak of the mouse embryo. *Nature* **432**, 625–630 (2004).
7. D'Souza, S. L., Elefanti, A. G. & Keller, G. SCL/Tal-1 is essential for hematopoietic commitment of the hemangioblast but not for its development. *Blood* **105**, 3862–3870 (2005).
8. Kimmel, C. B., Warga, R. M. & Schilling, T. F. Origin and organization of the zebrafish fate map. *Development* **108**, 581–594 (1990).
9. Warga, R. M. & Nusslein-Volhard, C. Origin and development of the zebrafish endoderm. *Development* **126**, 827–838 (1999).
10. Kinder, S. J. et al. The orderly allocation of mesodermal cells to the extraembryonic structures and the anteroposterior axis during gastrulation of the mouse embryo. *Development* **126**, 4691–4701 (1999).
11. Parker, L. H., Zon, L. & Stainier, D. Y. R. in *Methods in Cell Biology, The Zebrafish* (eds Detrich, W., Zon, L. & Westerfield, M.) 313–336 (Academic, San Diego, California, 1999).
12. Gering, M. & Patient, R. Hedgehog signaling is required for adult blood stem cell formation in zebrafish embryos. *Dev. Cell* **8**, 389–400 (2005).
13. Westerfield, M. *The Zebrafish Book* (Univ. of Oregon Press, Eugene, Oregon, 1993).
14. Jin, S. W., Beis, D., Mitchell, T., Chen, J. N. & Stainier, D. Y. Cellular and molecular analyses of vascular tube and lumen formation in zebrafish. *Development* **132**, 5199–5209 (2005).
15. Traver, D. et al. Transplantation and *in vivo* imaging of multilineage engraftment in zebrafish bloodless mutants. *Nature Immunol.* **4**, 1238–1246 (2003).
16. Horne-Badovinac, S., Rebagliati, M. & Stainier, D. Y. A cellular framework for gut-looping morphogenesis in zebrafish. *Science* **302**, 662–665 (2003).
17. Kozlowski, D. J., Murakami, T., Ho, R. K. & Weinberg, E. S. Regional cell movement and tissue patterning in the zebrafish embryo revealed by fate mapping with caged fluorescein. *Biochem. Cell Biol.* **75**, 551–562 (1997).
18. Keegan, B. R., Meyer, D. & Yelon, D. Organization of cardiac chamber progenitors in the zebrafish blastula. *Development* **131**, 3081–3091 (2004).

Acknowledgements We thank members of the Stainier and Martin laboratories for discussion and critical reading of the manuscript; D. Traver and L. Zon for the *Tg(gata1:DsRed)* line; and D. Yelon for advice with caged fluorescein. This study was supported by an NSF Pre-doctoral Fellowship (K.M.V.); an American Heart Association Post-doctoral Fellowship (S.-W.J.); the Packard foundation (D.Y.R.S.); and the National Institutes of Health (G.R.M. and D.Y.R.S.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.Y.R.S. (didier_stainier@biochem.ucsf.edu).

LETTERS

Topical drug rescue strategy and skin protection based on the role of *Mc1r* in UV-induced tanning

John A. D'Orazio^{1,2,3}, Tetsuji Nobuhisa^{1,2}, Rutao Cui^{1,2}, Michelle Arya^{1,2}, Malinda Spry³, Kazumasa Wakamatsu⁴, Vivien Igras^{1,2}, Takahiro Kunisada⁵, Scott R. Granter^{1,6}, Emi K. Nishimura^{1,2,7}, Shosuke Ito⁴ & David E. Fisher^{1,2}

Ultraviolet-light (UV)-induced tanning is defective in numerous 'fair-skinned' individuals, many of whom contain functional disruption of the melanocortin 1 receptor (MC1R)^{1–3}. Although this suggested a critical role for the MC1R ligand melanocyte stimulating hormone (MSH) in this response, a genetically controlled system has been lacking in which to determine the precise role of MSH–MC1R. Here we show that ultraviolet light potently induces expression of MSH in keratinocytes, but fails to stimulate pigmentation in the absence of functional MC1R in red/blonde-haired *Mc1r*^{e/e} mice. However, pigmentation could be rescued by topical application of the cyclic AMP agonist forskolin, without the need for ultraviolet light, demonstrating that the pigmentation machinery is available despite the absence of functional MC1R. This chemically induced pigmentation was protective against ultraviolet-light-induced cutaneous DNA damage and tumorigenesis when tested in the cancer-prone, xeroderma-pigmentosum-complementation-group-C-deficient genetic background. These data emphasize the essential role of intercellular MSH signalling in the tanning response, and suggest a clinical strategy for topical small-molecule manipulation of pigmentation.

Fair-skinned individuals have an increased incidence of skin cancer and often exhibit weak tanning responses⁴. Although multiple signalling pathways affect melanin production^{5,6}, 'fair' pigmentation in humans is largely the result of sequence variants in *MC1R*—encoding the receptor for MSH^{1,7,8}—that generate weak ligand-induced cAMP responses². Deficient tanning in *MC1R* variant individuals is consistent with a critical role for MSH/cAMP in this response³, but some studies have indicated that melanocyte-directed DNA damage might mediate UV-induced pigmentation^{9,10}. We compared UV tanning responses of wild-type C57BL/6 mice with an intact MSH pathway (*Mc1r*^{E/E}) to those of animals possessing an inactivating mutation of the MSH receptor (*Mc1r*^{e/e}, formerly known as *extension*), the homologue of a gene implicated in having fair skin in humans. Whereas UV-induced hyperpigmentation was grossly and microscopically observed in the ears of control mice, *Mc1r*^{e/e} (blonde-haired/pheomelanotic) mice lacked detectable pigmentation changes (Fig. 1a) despite comparable melanocyte numbers (Supplementary Fig. 1a). Murine pinnae (external ears) resemble human skin because of the presence of epidermal melanocytes that are lacking in truncal/fur skin.

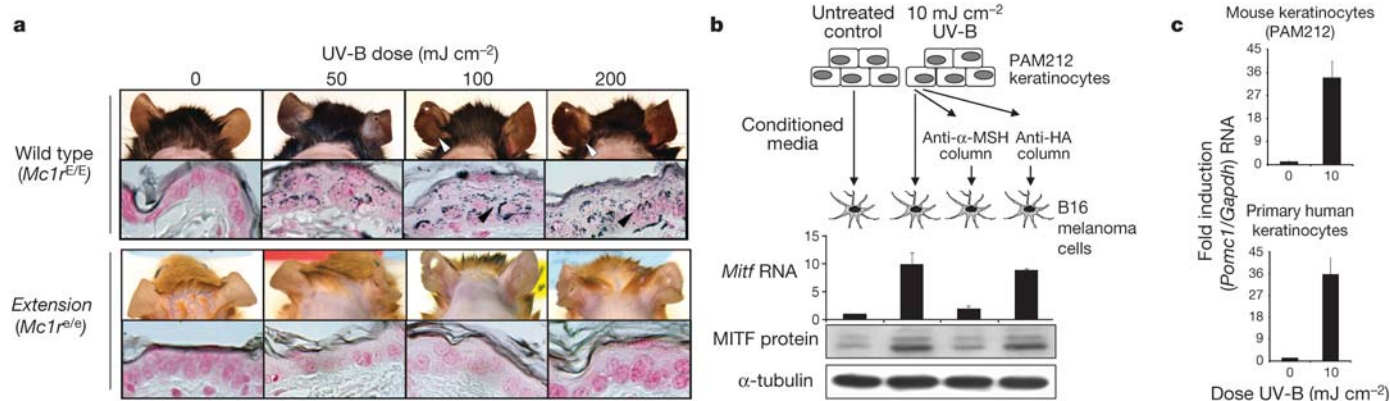


Figure 1 | UV-induced tanning requires an intact MSH pathway. **a**, C57BL/6 mice, either wild-type (*Mc1r*^{E/E}) or mutant (*Mc1r*^{e/e}) at the MSH receptor, were treated five days per week with daily UV-B delivered by a double bank of UV-B lamps (peak emission at 302 nm) for one month. Upper rows show UV-induced ear skin darkening, with corresponding Fontana–Masson-stained skin sections immediately below to show melanin accumulation (black deposits, $\times 400$ magnification). Note the UV-induced skin darkening (white arrows) and melanin accumulation (black arrows) in *Mc1r*^{E/E}, but not in *Mc1r*^{e/e} animals. Although *Mc1r*^{e/e} skin seems unpigmented, *Dct-LacZ*-tagged transgenic analysis reveals that it contains comparable melanocyte numbers

(Supplementary Fig. 1a). This experiment was repeated with similar results. **b**, *Mitf* induction by qPCR and by western analysis of B16 melanoma cells incubated (6 h) with 24-h conditioned supernatants from mouse keratinocytes either untreated (first lane) or irradiated with 10 mJ cm⁻² UV-B (remaining three lanes). Anti-MSH affinity chromatography (but not control anti-HA) abrogated *Mitf* induction by UV-conditioned media. **c**, qPCR-based detection of *Pomc1* mRNA 6 h after UV irradiation of PAM212 mouse keratinocytes or primary human keratinocytes. All PCR data are reported as fold induction by UV, normalized to *Gapdh* (control); samples were done in triplicate, and the standard deviations (s.d.) between samples are shown.

¹Melanoma Program and ²Department of Pediatric Hematology/Oncology, Dana-Farber Cancer Institute & Children's Hospital, 44 Binney Street, Boston, Massachusetts 02115, USA. ³Department of Pediatrics, Markey Cancer Center and the Graduate Center for Toxicology, University of Kentucky College of Medicine, Lexington, Kentucky 40536, USA. ⁴Department of Chemistry, Fujita Health University, School of Health Sciences, Toyoake, Aichi 470-1192, Japan. ⁵Department of Tissue and Organ Development, Gifu University, Graduate School of Medicine, 1-1 Yanagido, Gifu 501, Japan. ⁶Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. ⁷Department of Stem Cell Medicine, Kanazawa University, Cancer Research Institute, 13-1 Takaramachi, Kanazawa 920-0934, Japan.

To examine whether a non-cell-autonomous pathway may mediate the UV pigmentation effect, PAM212 keratinocytes were exposed to UV *in vitro* and culture supernatants were incubated with B16 melanoma cells to test for induction of the microphthalmia transcription factor, MITF¹¹. Media from UV-irradiated (versus non-irradiated) keratinocytes induced MITF messenger RNA and protein in melanoma cells (Fig. 1b). Absorption of conditioned media using anti- α -MSH affinity chromatography removed most of the stimulatory activity, which was not removed with anti-(haemagglutinin) HA. Comparable conditioned media from UV-irradiated murine or primary human melanocytes lacked this activity (Supplementary Fig. 1b). Direct measurements revealed >30-fold induction of *Msh* (otherwise known as pro-opiomelanocortin alpha (*Pomc1*)) expression by UV in PAM212 mouse keratinocytes and primary human keratinocytes (Fig. 1c), whereas <4-fold induction was observed in melanocytes (Supplementary Fig. 1c). These observations corroborate previous findings that UV radiation activates expression of *Msh* in keratinocytes¹². Although melanocytes showed relatively weak UV-induction of *Pomc1* expression, they nonetheless did express some *Pomc1* with or without UV exposure, as previously described^{13,14}, indicating that autocrine signalling could still contribute to MC1R activity.

The UV-unresponsiveness of *Mc1r* variants could, in principle, arise from an irreversible block to eumelanin synthesis occurring at some developmental stage. If a cAMP agonist could restore pigmentation in this setting, then the inability of UV to induce pigmentation would be more rigorously linked to inactivity of the cAMP pathway within the *Mc1r*^{e/e} genetic background. To this end, we used congenic C57BL/6 mice harbouring a transgene in which the keratin 14 promoter drives constitutive expression of stem cell factor (*Scf*; also known as *Kitl*) in the epidermis¹⁵: *K14-Scf* mimics the human

pattern of keratinocyte *SCF* expression as well as epidermal melanocyte homing (the presence of melanocytes in the basal layer of the epidermis), independent of MC1R status (Supplementary Fig. 1d). *K14-Scf* transgenic ('humanized') mice, homozygous for *Mc1r*^{e/e}, exhibited fair/pink skin with high pheomelanin and low eumelanin content (Supplementary Fig. 2a–c).

Just as in non-transgenic *Mc1r*^{e/e} mice, we observed no measurable UV-induced melanization in *K14-Scf*-transgenic, *Mc1r*^{e/e} mice (Fig. 2a; Supplementary Fig. 2d). When forskolin, a cell-permeable diterpenoid that activates adenyl cyclase¹⁶, was topically applied to *Mc1r*^{e/e}; *K14-Scf* (pheomelanotic) mice, significant melanization was observed, both when applied throughout (Fig. 2b; Supplementary Fig. 2d) and when applied only to the rump area (Supplementary Fig. 3c). Daily topical forskolin caused progressive and robust darkening (Fig. 2a, Supplementary Fig. 3a), and we conclude that forskolin-induced eumelanization required epidermal melanocytes, because it was not observed in the truncal skin of mice lacking the *K14-Scf* transgene (Supplementary Fig. 3b). Skin darkening by forskolin was associated with dose-dependent accumulation of melanin in the epidermis (Fig. 2c), was progressive over several weeks, and was reversible with a half-life of approximately 2 weeks (Supplementary Fig. 3a). Melanin quantification¹⁷ revealed a >20-fold increase in eumelanin (Fig. 2d) in treated, depilated skin, together with a shift in the eumelanin:pheomelanin ratio from 0.09 ± 0.1 to 0.9 ± 0.5 .

Hair/fur colour was not markedly altered during the time courses examined here. Both crude preparations of forskolin (root extract of *Plectranthus barbatus*, otherwise known as *Coleus forskohlii*)¹⁸ and chemically pure forskolin produced significant melanization (Supplementary Fig. 4a). Importantly, forskolin treatment mimicked melanization to the degree of exhibiting 'nuclear capping', in which

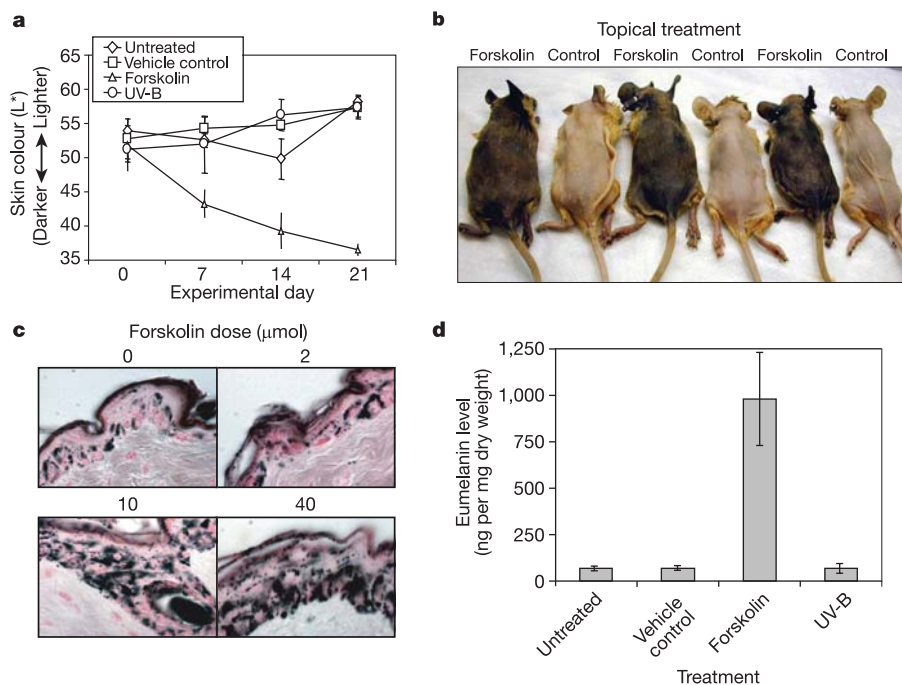


Figure 2 | Forskolin, but not UV, rescues eumelanin production in mice with defective MSH signalling. **a**, Reflective colorimetry measurements (CIE L* white–black colour axis)²⁹ of skin darkening of *Mc1r*^{e/e}; *K14-Scf*-transgenic, depilated C57BL/6 mice either topically treated with vehicle control (70% ethanol, 30% propylene glycol), irradiated with 200 mJ cm⁻² UV-B, topically treated with 400 μ l of *P. barbatus* root extract (80 μ mol forskolin) to the dorsal skin surface for 21 days, or untreated. Lower numbers represent darker skin tones. Skin colour measurements were done in triplicate (different anatomic locations within the treated skin field) and s.d. between

measurements is shown; this experiment was repeated many times with similar results. **b**, Side-by-side photographs of vehicle-control- or forskolin-treated *Mc1r*^{e/e}; *K14-Scf* animals treated as described in **a**. **c**, Fontana–Masson (eumelanin)-stained skin sections of animals treated for 21 days with the indicated dose of forskolin derived from *P. barbatus* extract ($\times 630$ magnification) as described in **a**. **d**, Eumelanin levels ($\pm$ s.d.) of whole, depilated skin from *Mc1r*^{e/e}; *K14-Scf* mice treated as described in **a** (ref. 17). Melanin quantification measurements were done in triplicate.

melanosomes exported to keratinocytes align in discrete structures that putatively provide solar shielding to keratinocyte nuclei (Fig. 3a, see asterisks)¹⁹. We conclude that forskolin-induced melanization did not require ectopic SCF because it was also induced on pinnae of non-*K14-Scf*-transgenic (C57BL/6) mice (Supplementary Fig. 4c). Collectively, these data demonstrate strong rescue by forskolin of eumelanin production in MC1R-deficient melanocytes, thus indicating that the inability of UV to trigger eumelanization might not be caused by irreversible inactivation of pigmentation machinery.

We considered that small-molecule-induced pigmentation might be UV-protective, so we examined formation of sunburn cells (apoptotic epidermal keratinocytes²⁰) 24 h after a single 200 mJ cm⁻² dose of UV-B (arrows, Fig. 3a). Whereas vehicle-control-treated *Mc1r*^{+/+} mice were very UV-sensitive, forskolin pre-treatment produced nearly the same degree of UV protection as found in genetically black-skinned (*Mc1r*^{E/E}) mice (Fig. 3b, c). Tyrosinase mutants (albino) were not protected by forskolin (Fig. 3b, c), suggesting that forskolin-induced keratinocyte survival was mediated by pigmentation, rather than pigment-independent effects. Fluorescence staining for cyclobutane dimers—the mutagenic and most abundant DNA lesion caused by UV-B—revealed dose-dependent nuclear staining at 20 or 50 mJ cm⁻² UV-B in *Mc1r*^{+/+} and albino (*Tyr*^{c2j/c2j}) skin, but little in *Mc1r*^{E/E} or forskolin-treated *Mc1r*^{+/+} mice (Fig. 3c). Forskolin-induced melanization was nearly as protective as *Mc1r*^{E/E} (genetically black) epidermis.

To test whether forskolin treatment might protect against UV

carcinogenesis, xeroderma-pigmentosum-complementation-group-C-deficient mice (*Xpc*^{-/-})²¹ were crossed to the fair-skinned *Mc1r*^{+/+}, *K14-Scf* (C57BL/6 background) mice and subjected to either forskolin-containing or vehicle-control topical treatments for four weeks, before daily exposure to 250 mJ cm⁻² UV-B (along with continued topical treatments) for 20 weeks—a UV dose approximating 1–2 h of ambient midday sun exposure at sea level in Florida, during July (<http://www.srrb.noaa.gov/UV/>). This low-dose chronic UV schedule was chosen for its propensity to induce keratinocyte neoplasms²¹. Vehicle control (non-darkened) mice exhibited gross and histologic evidence of sun damage, including failure to thrive (diminished weight-gain; Supplementary Fig. 5a, b), profound epidermal thickening, and inflammation and scarring (Supplementary Fig. 6a, b). All such damage was significantly prevented by forskolin pretreatment. Following chronic UV, eleven neoplasms developed in the nine vehicle-control, irradiated mice within nine weeks of cessation of UV radiation (Fig. 4a, b), with two of the nine mice developing multiple tumours. As anticipated (by dose and schedule regimen), the majority were squamous cell carcinomas (Fig. 4b). Forskolin pretreatment was significantly protective in delaying onset of UV-induced tumours (median of four weeks in controls versus 25 weeks in forskolin-treated; *P* < 0.001). In all, vehicle-control-treated, irradiated mice developed 11 tumours whereas forskolin-treated, irradiated mice developed six tumours. Given the extreme UV sensitivity of homozygous *Xpc* deficiency, coupled to the high chronic UV doses used, these data suggest a

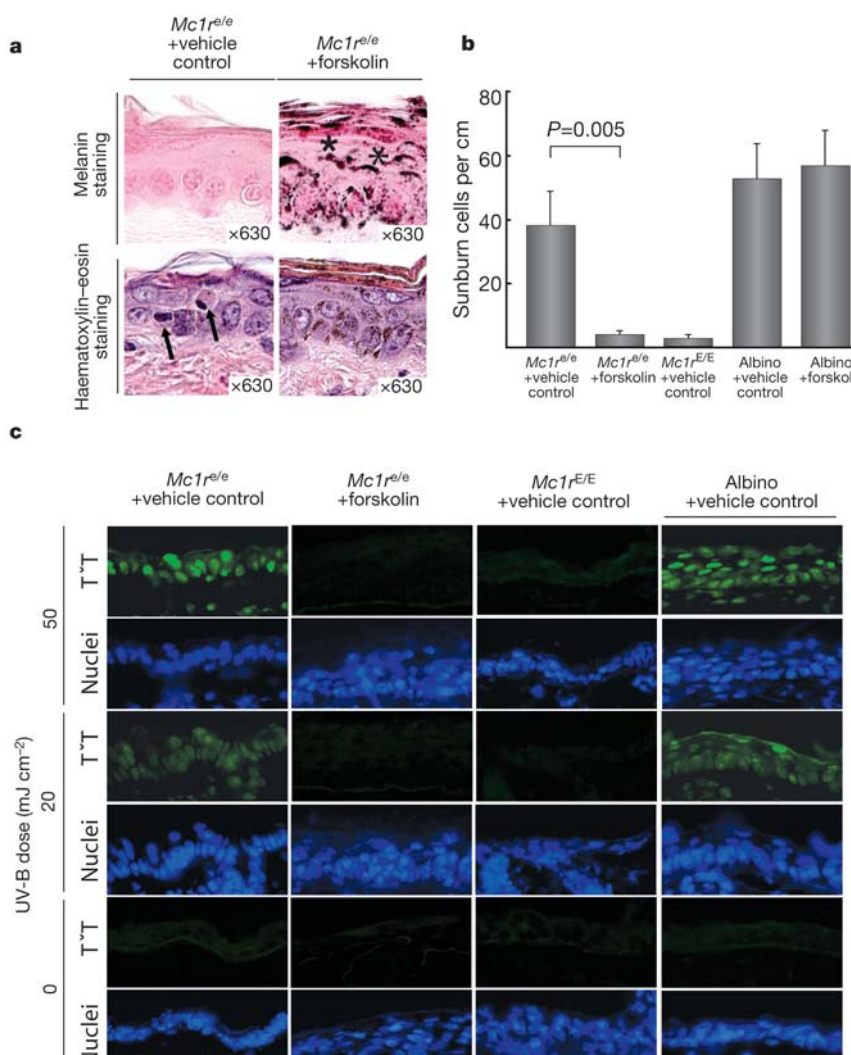


Figure 3 | Forskolin-induced melanin deposition protects against UV-mediated cutaneous damage.

Mc1r^{+/+}, *K14-Scf*, depilated animals were pretreated for 15 days with either vehicle control or *P. barbatum* root extract (80 μmol forskolin) before exposure to 200 mJ cm⁻² UV-B. **a**, Melanin distribution in the skin 24 h after UV exposure as determined by Fontana–Masson (upper panels) or haematoxylin–eosin staining (lower panels). Note the greatly enhanced deposition of epidermal melanin (black staining) in forskolin-treated animals in a keratinocyte nucleus-capping pattern (see asterisks) and the presence of ‘sunburn cells’, which represent apoptotic keratinocytes in UV-irradiated animals not pre-treated with forskolin (arrows). **b**, Sunburn cell quantification in UV-exposed depilated animals treated as described in **a**. Shown are means ± s.d. of at least three samples per condition. Forskolin pre-treatment of fair-skinned animals yielded as much protection as *Mc1r*^{E/E} (eumelanotic) animals. Amelanotic (albino) animals had high levels of UV-induced apoptosis regardless of whether they were treated with forskolin before UV exposure, suggesting that forskolin’s protective effect is pigmentation-dependent. **c**, UV-induced thymine dimer formation in depilated skin harvested ten minutes after UV irradiation from *Mc1r*^{+/+}, *Mc1r*^{E/E} or *Tyr*^{c2j/c2j} (albino); *K14-Scf* mice pretreated with either vehicle control or *P. barbatum* root extract (80 μmol forskolin). DAPI stain highlights epidermal nuclei (blue), as does thymine-dimer-directed immunofluorescence (T~T, green) in DNA-damaged skin in a dose-dependent fashion. Non-fluorescent (immunohistochemical) staining for pyrimidine dimers revealed the same protection, ruling out potential fluorescence quenching by melanin (Supplementary Fig. 4b). This experiment was repeated at least once with similar results.

significant protection afforded by topical rescue of MC1R functional deficiency.

A requirement for MSH/MC1R signalling in UV-induced pigmentation is plausible, given previous studies of pheomelanin production^{2,3}, agouti signalling protein (a natural MC1R antagonist)²², and UV induction of *Msh* expression in keratinocytes¹². UV also upregulated *MSH* in melanocytes (as previously reported^{13,14}), but it did so to a much smaller degree than in keratinocytes. These studies therefore cannot firmly discriminate the relative roles of keratinocytes versus melanocyte MSH (paracrine versus autocrine) in the observed UV response, and indeed it is possible that both are contributory. Furthermore, the observed essential role of MC1R signalling in the tanning response does not diminish the potential contributions of other pathways. Thus, although forskolin was singularly capable of restoring dramatic melanization, other signalling factors (endothelin-1, β -FGF, NO, SCF, p38, USF and others) have been implicated in the UV adaptive tanning response^{23–28} and it will be valuable to determine their relative contributions.

It will also be important to understand the pathway through which UV induces *Pomc1/Msh* expression in keratinocytes, because it is possible that alterations in *Msh* induction by UV could contribute to cancer risk (for example, in dark-haired, fair-skinned people). Previous work with topical cAMP agonists in other animal models showed that forskolin did not darken light-skinned swine, but did promote histologically identifiable melanin accumulation in the epidermis¹⁸ (although this was not studied in a UV-resistant *Mc1r* variant background). The greater robustness of the forskolin response in mice may involve a thinner epidermis or genetic features that are ill-defined in swine. Clearly, small-molecule delivery would

need to be optimized before pigmentation in the absence of sun could be successful in humans.

Whereas rescue of eumelanization was observed with topical forskolin, the cAMP agonist effect was not explicitly targeted to melanocytes within the skin. *In vitro* studies showed no measurable melanocytic stimulation from cultured supernatants of forskolin-treated keratinocytes (data not shown). Still, it is theoretically possible that part of the rescue may have involved non-cAMP or pigment-independent effects, or even potential impurities within the small-molecule preparation, although forskolin could not rescue keratinocyte apoptosis in UV-irradiated albinos. The topical rescue of eumelanization confirms the 'availability' of the pigmentation machinery in adults, given appropriate signals. It remains to be seen whether topical melanization will be achievable in man, and whether it would afford measurable protection against UV skin damage and cancer.

METHODS

See Supplementary Information for detailed methods.

Mice. C57BL/6J *Mc1r*^{E/E}; *Tyr*^{+/+} mice, C57BL/6J *Mc1r*^{e/e} mice and C57BL/6J *Mc1r*^{E/E}; *Tyr*^{c2j/c2j} mice were crossed with *K14-Scf* transgenic mice¹⁵, dopa-chrome tautomerase (*Dct*)-*LacZ* transgenic mice or *Xpc*-null animals.

Melanization experiments. Depilated animals were exposed to either UV-B with UV-B lamps (UV Products), topical forskolin or vehicle control. Skin reflective colorimetry measurements were assessed with a CR-400 Colorimeter (Minolta) and the degree of melanization was calculated²⁹. Skin biopsies were harvested and processed for Fontana–Masson (eumelanin) staining and melanin quantification¹⁷.

Histology and immunohistochemistry. Skin biopsies were fixed and paraffin-sectioned (haematoxylin–eosin or Fontana–Masson staining) or frozen-sectioned (immunohistochemistry and β -galactosidase staining)³⁰. Immunohistochemistry used an anti-thymine-dimer, monoclonal antibody (Kamiya Biomedical) followed by Alexa-Fluor-488-conjugated, anti-mouse IgG donkey antibody (Molecular Probes).

Quantitative polymerase chain reaction and western blotting. Melanocyte or keratinocyte cells were exposed to UV radiation, anti-MSH or treated cell supernatants as indicated. MSH was neutralized from cell supernatants using anti-MSH (Sigma). MITF detection by western blotting employed the C5 monoclonal antibody¹¹. RNA was extracted and *Mitf* mRNA expression was quantified by quantitative Taqman polymerase chain reaction (qPCR) using QuantiTect Probe reverse transcription (RT)–PCR kits (Qiagen).

Tumour formation and chronic UV protection experiments. C57BL/6 *Mc1r*^{e/e}; *K14-Scf*; *Xpc*^{−/−} animals were irradiated (250 mJ cm^{−2} day^{−1}) beginning at seven weeks of age over the course of 20 weeks, after one month's topical pretreatment with either forskolin or vehicle control. Animals were weighed and skin samples were harvested after 16 weeks of irradiation. Animals were then monitored for tumour surveillance over the next 52 weeks. Lesions that were grossly identified as tumours were biopsied and examined.

Received 28 March; accepted 14 July 2006.

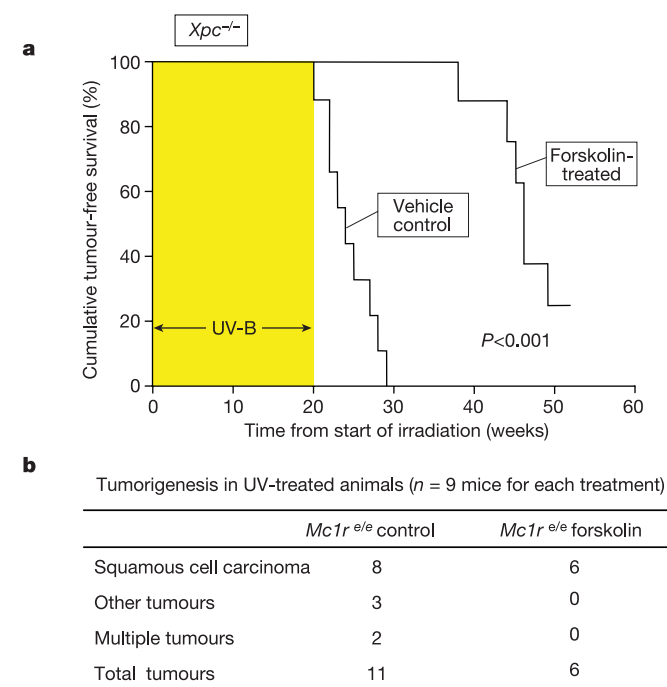


Figure 4 | Protective effect of topical forskolin against chronic UV damage. *K14-Scf*-transgenic, *Mc1r*^{e/e}, depilated animals homozygously deficient for the *Xpc* gene²¹ were pretreated with either vehicle control or topical forskolin and exposed to daily low-dose UV-B for 16–20 weeks. **a**, Kaplan–Meier analysis of tumour incidence in irradiated, vehicle-control-treated versus forskolin-treated, irradiated *Mc1r*^{e/e}; *K14-Scf*; *Xpc*^{−/−} mice. The yellow bar indicates duration of daily UV exposure. **b**, Tumour subtypes and incidence in vehicle-control-treated versus forskolin-treated, irradiated *Mc1r*^{e/e}; *K14-Scf*; *Xpc*^{−/−} animals. Some vehicle-control-treated, irradiated animals developed multiple distinct tumours, whereas none of the forskolin-treated animals developed more than one tumour.

- Mountjoy, K. G., Robbins, L. S., Mortrud, M. T. & Cone, R. D. The cloning of a family of genes that encode the melanocortin receptors. *Science* **257**, 1248–1251 (1992).
- Valverde, P., Healy, E., Jackson, I., Rees, J. L. & Thody, A. J. Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. *Nature Genet.* **11**, 328–330 (1995).
- Rouzaud, F., Kadekaro, A. L., Abdel-Malek, Z. A. & Hearing, V. J. MC1R and the response of melanocytes to ultraviolet radiation. *Mutat. Res.* **571**, 133–152 (2005).
- Sturm, R. A. et al. Genetic association and cellular function of MC1R variant alleles in human pigmentation. *Ann. NY Acad. Sci.* **994**, 348–358 (2003).
- Van Raamsdonk, C. D., Fitch, K. R., Fuchs, H., de Angelis, M. H. & Barsh, G. S. Effects of G-protein mutations on skin color. *Nature Genet.* **36**, 961–968 (2004).
- Lamason, R. L. et al. SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* **310**, 1782–1786 (2005).
- Naysmith, L. et al. Quantitative measures of the effect of the melanocortin 1 receptor on human pigmentation status. *J. Invest. Dermatol.* **122**, 423–428 (2004).
- Rees, J. L. The genetics of sun sensitivity in humans. *Am. J. Hum. Genet.* **75**, 739–751 (2004).
- Eller, M. S., Yaar, M. & Gilchrist, B. A. DNA damage and melanogenesis. *Nature* **372**, 413–414 (1994).

10. Eller, M. S., Ostrom, K. & Gilchrist, B. A. DNA damage enhances melanogenesis. *Proc. Natl Acad. Sci. USA* **93**, 1087–1092 (1996).
11. Price, E. R. *et al.* α -melanocyte-stimulating hormone signaling regulates expression of *microphthalmia*, a gene deficient in Waardenburg syndrome. *J. Biol. Chem.* **273**, 33042–33047 (1998).
12. Wintzen, M., Yaar, M., Burbach, J. P. & Gilchrist, B. A. Proopiomelanocortin gene product regulation in keratinocytes. *J. Invest. Dermatol.* **106**, 673–678 (1996).
13. Corre, S. *et al.* UV-induced expression of key component of the tanning process, the *POMC* and *MC1R* genes, is dependent on the p-38-activated upstream stimulating factor-1 (USF-1). *J. Biol. Chem.* **279**, 51226–51233 (2004).
14. Chakraborty, A. K. *et al.* Production and release of proopiomelanocortin (POMC) derived peptides by human melanocytes and keratinocytes in culture: regulation by ultraviolet B. *Biochim. Biophys. Acta* **1313**, 130–138 (1996).
15. Kunisada, T. *et al.* Murine cutaneous mastocytosis and epidermal melanocytosis induced by keratinocyte expression of transgenic stem cell factor. *J. Exp. Med.* **187**, 1565–1573 (1998).
16. Seamon, K. B. & Daly, J. W. Forskolin: a unique diterpene activator of cyclic AMP-generating systems. *J. Cyclic Nucleotide Res.* **7**, 201–224 (1981).
17. Wakamatsu, K. & Ito, S. Advanced chemical methods in melanin determination. *Pigment Cell Res.* **15**, 174–183 (2002).
18. Lin, C. B. *et al.* Modulation of microphthalmia-associated transcription factor gene expression alters skin pigmentation. *J. Invest. Dermatol.* **119**, 1330–1340 (2002).
19. Kobayashi, N. *et al.* Supranuclear melanin caps reduce ultraviolet induced DNA photoproducts in human epidermis. *J. Invest. Dermatol.* **110**, 806–810 (1998).
20. Bayerl, C., Taake, S., Moll, I. & Jung, E. G. Characterization of sunburn cells after exposure to ultraviolet light. *Photodermatol. Photoimmunol. Photomed.* **11**, 149–154 (1995).
21. Sands, A. T., Abuin, A., Sanchez, A., Conti, C. J. & Bradley, A. High susceptibility to ultraviolet-induced carcinogenesis in mice lacking *XPC*. *Nature* **377**, 162–165 (1995).
22. Sakai, C. *et al.* Modulation of murine melanocyte function *in vitro* by agouti signal protein. *EMBO J.* **16**, 3544–3552 (1997).
23. Hirobe, T. Basic fibroblast growth factor stimulates the sustained proliferation of mouse epidermal melanoblasts in a serum-free medium in the presence of dibutyryl cyclic AMP and keratinocytes. *Development* **114**, 435–445 (1992).
24. Cook, A. L. *et al.* Human melanoblasts in culture: expression of *BRN2* and synergistic regulation by fibroblast growth factor-2, stem cell factor, and endothelin-3. *J. Invest. Dermatol.* **121**, 1150–1159 (2003).
25. Lassalle, M. W. *et al.* Effects of melanogenesis-inducing nitric oxide and histamine on the production of eumelanin and pheomelanin in cultured human melanocytes. *Pigment Cell Res.* **16**, 81–84 (2003).
26. Grichnik, J. M., Burch, J. A., Burchette, J. & Shea, C. R. The SCF/KIT pathway plays a critical role in the control of normal human melanocyte homeostasis. *J. Invest. Dermatol.* **111**, 233–238 (1998).
27. Kadekaro, A. L. *et al.* α -Melanocortin and endothelin-1 activate antiapoptotic pathways and reduce DNA damage in human melanocytes. *Cancer Res.* **65**, 4292–4299 (2005).
28. Galibert, M. D., Carreira, S. & Goding, C. R. The *Usf-1* transcription factor is a novel target for the stress-responsive p38 kinase and mediates UV-induced *Tyrosinase* expression. *EMBO J.* **20**, 5022–5031 (2001).
29. Park, S. B., Suh, D. H. & Youn, J. I. A long-term time course of colorimetric evaluation of ultraviolet light-induced skin reactions. *Clin. Exp. Dermatol.* **24**, 315–320 (1999).
30. Nishimura, E. K., Granter, S. R. & Fisher, D. E. Mechanisms of hair graying: incomplete melanocyte stem cell maintenance in the niche. *Science* **307**, 720–724 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Tsay, J. Du, H. Widlund, M. Seiberg, I. Davis and A. Wagner for discussions and help with technical aspects of the studies; I. Jackson for *Dct-LacZ* mice; S. Yuspa for PAM212 cells; and D. Bennett for Melan-C cells. We also thank the University of Kentucky's Teaching and Academic Support Center (TASC) for help with figure preparation. This work was supported by grants from the NIH (D.E.F.) and the Doris Duke Charitable Foundation. S.I. and K.W. were supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports and Technology of Japan. D.E.F. is the Jan and Charles Nirenberg Fellow in Pediatric Oncology at the Dana-Farber Cancer Institute, and a Doris Duke Distinguished Clinical Investigator.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.E.F. (david_fisher@dfci.harvard.edu).

Calcineurin/NFAT signalling regulates pancreatic β -cell growth and function

Jeremy J. Heit¹, Åsa A. Apelqvist^{1†}, Xueying Gu¹, Monte M. Winslow², Joel R. Neilson³, Gerald R. Crabtree^{1,4} & Seung K. Kim^{1,5}

The growth and function of organs such as pancreatic islets adapt to meet physiological challenges and maintain metabolic balance, but the mechanisms controlling these facultative responses are unclear^{1,2}. Diabetes in patients treated with calcineurin inhibitors such as cyclosporin A indicates that calcineurin/nuclear factor of activated T-cells (NFAT) signalling might control adaptive islet responses³, but the roles of this pathway in β -cells *in vivo* are not understood. Here we show that mice with a β -cell-specific deletion of the calcineurin phosphatase regulatory subunit, calcineurin b1 (*Cnb1*), develop age-dependent diabetes characterized by decreased β -cell proliferation and mass, reduced pancreatic insulin content and hypoinsulinaemia. Moreover, β -cells lacking *Cnb1* have a reduced expression of established regulators of β -cell proliferation^{1,4,5}. Conditional expression of active NFATc1 in *Cnb1*-deficient β -cells rescues these defects and prevents diabetes. In normal adult β -cells, conditional NFAT activation promotes the expression of cell-cycle regulators and increases β -cell proliferation and mass, resulting in hyperinsulinaemia. Conditional NFAT activation also induces the expression of genes critical for β -cell endocrine function, including all six genes mutated in hereditary forms of monogenic type 2 diabetes. Thus, calcineurin/NFAT signalling regulates multiple factors that control β -cell growth and hallmark β -cell functions, revealing unique models for the pathogenesis and therapy of diabetes.

Physiological and developmental cues in cells such as lymphocytes, myocytes and neurons stimulate intracellular Ca^{2+} transients, leading to dephosphorylation and nuclear translocation of the cytoplasmic subunits of NFAT transcription complexes (NFATc)^{6–8}. β -cell mitogens and secretion stimulators such as glucose, insulin and glucagon-like peptide-1 (GLP-1) promote increased intracellular Ca^{2+} concentrations, indicating that calcineurin/NFAT signalling might regulate β -cell proliferation and secretion¹. Consistent with this possibility is the observation that all four NFATc proteins (NFATc1–4) are expressed in β -cells as well as in other hormone-producing islet cells and a subset of acinar cells (Fig. 1a–d and Supplementary Fig. 1). *Cnb1* regulates the activity of the calcineurin phosphatase complex and is required for the activation of all NFATc proteins^{6,7}. We therefore constructed mice with specific inactivation of *Cnb1* in β -cells. Mice with *Cnb1* floxed (*Cnb1*^f) and null (*Cnb1*^Δ) alleles^{8,9} or with a transgene expressing Cre recombinase from the rat insulin promoter (RIP; ref. 10) were used to generate RIP-Cre; *Cnb1*^{Δ/f} (β Cnb1KO) mice. Littermate control islets expressed *Cnb1* messenger RNA (Fig. 1e), but β Cnb1KO islets lacked *Cnb1* mRNA (Fig. 1e), which is consistent with genomic *Cnb1*^f allele recombination in islets (Supplementary Fig. 2). In contrast, no RIP-Cre activity was detected in other tissues, including the brain (Supplementary Fig. 2). As expected, the accumulation of NFATc1 in β -cell nuclei was reduced

in β Cnb1KO mice, in contrast to the predominantly nuclear localization of NFATc1 in littermate controls (Fig. 1f, g).

To study the consequences of *Cnb1* inactivation in β Cnb1KO mice, we measured blood glucose levels during random feeding as these mice aged. Littermate controls (*Cnb1*^{+/f}, *Cnb1*^{Δ/f} and RIP-Cre; *Cnb1*^{+/f} mice, which were indistinguishable) maintained normoglycaemia, but β Cnb1KO mice had increasingly severe hyperglycaemia and developed overt diabetes mellitus by 10 weeks of age (Fig. 2a). Moreover, after acute glucose challenge by intraperitoneal injection of glucose, β Cnb1KO animals had significantly elevated blood glucose levels and prolonged hyperglycaemia in comparison with littermate controls (Fig. 2b). We did not detect impaired insulin sensitivity in β Cnb1KO mice (Supplementary Fig. 3). Thus, *Cnb1* in β -cells is essential for maintaining normoglycaemia and preventing diabetes mellitus.

In vitro studies indicated that NFATc proteins might stimulate the transcription of the gene encoding insulin^{11–14}; thus, we postulated that *Cnb1* deficiency might disrupt the expression of insulin, a cardinal β -cell product. Random-fed β Cnb1KO mice had a 50% decrease in serum insulin levels compared with littermates (Fig. 2c).

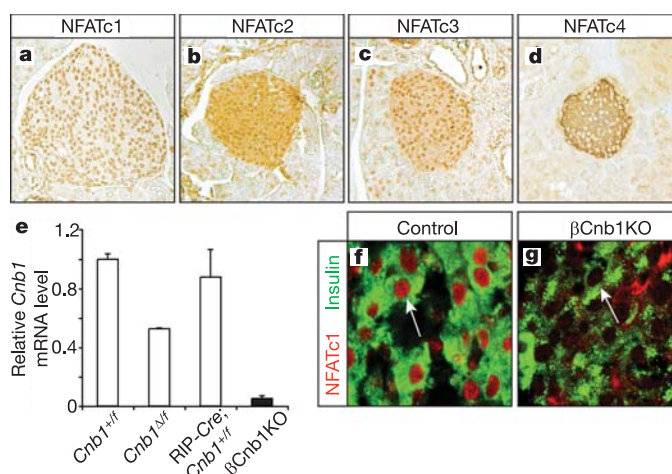


Figure 1 | NFATc proteins are expressed in β -cells, and *Cnb1* deletion results in NFATc1 mislocalization. a–d, Expression of NFATc proteins in β -cells detected with antibodies specific for NFATc1 (a), NFATc2 (b), NFATc3 (c) or NFATc4 (d). e, *Cnb1* RT-PCR in handpicked islets from 8-week-old mice with indicated genotypes. Results are presented as means $\pm$ s.e.m. f, g, NFATc1 immunostaining. Nuclear NFATc1 is readily detected in control β -cells (f, arrow), but only cytoplasmic NFATc1 is observed in β Cnb1KO β -cells (g, arrow).

¹Department of Developmental Biology, ²Program in Immunology, ³Department of Microbiology and Immunology, ⁴Department of Pathology and Howard Hughes Medical Institute, ⁵Department of Medicine (Oncology Division), Stanford University, Stanford, California 94305, USA. †Present address: MRC Centre for Developmental Neurobiology, King's College London, London WC2R 2LS, UK.

Consistent with this hypoinsulinaemia, pancreatic insulin production in β Cnb1KO mice was decreased by more than 85% in comparison with littermate controls at 12 weeks of age (Fig. 2d). In contrast, pancreatic glucagon content and α -cell number were not detectably altered in β Cnb1KO mice (Fig. 2e and data not shown). Coupled with evidence of reduced transcription of the insulin gene by islets (see below), these results indicate that impaired insulin production might have contributed to diabetes pathogenesis in β Cnb1KO mice.

To test whether reduced pancreatic insulin content in β Cnb1KO mice reflects decreased β -cell mass, we assessed the pancreata of β Cnb1KO and littermate control mice at 4, 8 and 12 weeks of age, a period when body mass and total pancreatic mass double (Fig. 2f). The mass of β -cells at 4 and 8 weeks of age was not significantly different between β Cnb1KO and littermate control mice ($P > 0.1$; Fig. 2g). However, by 12 weeks the β -cell mass in β Cnb1KO mice was decreased by 50% in comparison with littermate controls ($P < 0.002$; Fig. 2g), reflecting a combination of decreased islet size and islet number (Supplementary Fig. 4). In β Cnb1KO mice, incorporation of bromodeoxyuridine (BrdU) showed that β -cell proliferation was significantly reduced at 4, 8 and 12 weeks (Fig. 2h). We did not

detect differences in pancreatic apoptosis between β Cnb1KO and littermate controls, indicating that cell death might not have contributed to decreased β -cell mass in β Cnb1KO mice (Supplementary Fig. 5). Taken together, these results show that calcineurin is essential for postnatal β -cell proliferation and establishing appropriate β -cell mass. Impaired β -cell secretion could also contribute to hypoinsulinaemia in β Cnb1KO mice, but we did not obtain unequivocal evidence of secretion defects in β Cnb1KO islets (Supplementary Fig. 6).

To identify the basis for β -cell failure in β Cnb1KO mice, we measured gene expression in β -cells. Immunohistological staining of insulin in β -cells was decreased; this was consistent with measures of total pancreatic insulin content (Fig. 3a, f). Levels of Glut2, a membrane-spanning glucose transporter required for glucose sensing by β -cells¹⁵, were markedly lower in β Cnb1KO islets (Fig. 3b, g), which was consistent with the possibility of impaired insulin secretion in β Cnb1KO mice (Supplementary Fig. 6). Immunohistology also revealed decreased expression of β -cell transcription factors known to regulate the genes encoding insulin and Glut2, including MafA, Pdx1 and Beta2 (also called NeuroD1; Fig. 3c–e, h–j). Mutations in *Pdx1*, *Beta2*, *Hnf4 α* , *Hnf1 α* , *Hnf1 β* or the glucokinase gene (*Gck*) can cause monogenic forms of type 2 diabetes known as maturity onset diabetes of the young (MODY)¹⁶. To quantify changes in β Cnb1KO islet gene expression, we measured levels of mRNAs specifying Insulin1 (*Ins1*), Insulin2 (*Ins2*), *Glut2*, *MafA* and the six MODY factors by real-time reverse transcriptase polymerase chain reaction (RT-PCR). In agreement with our immunohistology, we found that the levels of islet mRNAs for *Ins1*, *Ins2*, *Glut2*, *MafA*, *Pdx1* and *Beta2* were significantly decreased compared with those in littermate controls (Fig. 3k). Decreased expression of these genes is consistent with the attenuated insulin synthesis and secretion observed in β Cnb1KO mice. Islet mRNA levels for *Hnf4 α* , *Hnf1 α* , *Hnf1 β* and *Gck* were not significantly altered in β Cnb1KO mice compared with controls (Fig. 3k). We next measured levels of mRNAs encoding β -cell-cycle regulators, including cyclin-dependent kinase 4 (*Cdk4*), cyclin D1 (*Ccnd1*), cyclin D2 (*Ccnd2*) and *c-myc*^{4,5,17–19}. Cyclin D2 mRNA levels were significantly decreased in β Cnb1KO islets. Decreased mRNA levels for *Cdk4*, cyclin D1 and *c-myc* were also detected, but these decreases did not achieve statistical significance (Fig. 3l). Taken together, these data indicate that diabetes in β Cnb1KO mice results from the disrupted expression of essential β -cell factors that regulate proliferation and insulin production.

Our results indicated that NFATc proteins might control the transcription of several essential β -cell genes. To test this hypothesis we used antibodies specific for NFATc1 in chromatin immunoprecipitation (ChIP) studies to detect direct NFATc1 association with potential NFAT-binding sites identified by bioinformatics analysis in these genes (Fig. 3m, n, and Supplementary Fig. 7). Endogenous NFATc1 levels in the MIN6 mouse β -cell line²⁰ were nearly undetectable, but NFATc1 levels increased after transfection with complementary DNA encoding NFATc1. After calcineurin/NFAT activation with 12-*O*-tetradecanoylphorbol-13-acetate (TPA) and ionomycin, we detected NFATc1 associated with *cis*-regulatory elements of *Ins1*, *Hnf4 α* , *Gck*, *Hnf1 α* , *Pdx1*, *Hnf1 β* , *Beta2*, *Glut2*, *Cdk4* and *Ccnd1* (Fig. 3m, n). NFATc1 binding to these promoter-proximal sequences was abolished by pretreatment with cyclosporin A, indicating that active NFATc1 function was assessed (Fig. 3m, n). Furthermore, NFATc1 did not detectably associate with promoter regions that lacked NFAT consensus sequences (Supplementary Fig. 8). In the context of transcriptional activation²¹, these data provide unique evidence that calcineurin/NFAT signalling directly controls the expression of cell-cycle regulators and all known MODY factors.

Overt diabetes in β Cnb1KO mice demonstrates that calcineurin is essential for age-dependent β -cell expansion and normal β -cell function, and our ChIP studies indicate that NFATc activity may

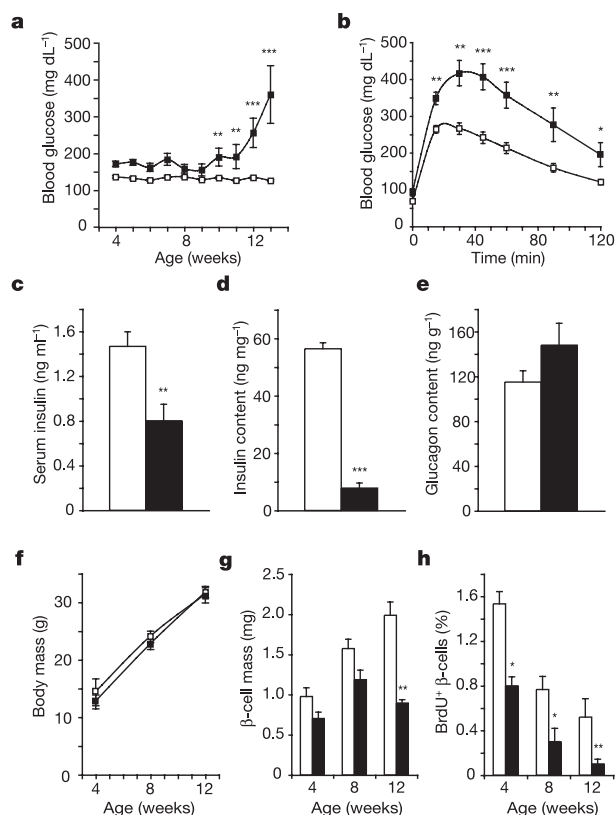


Figure 2 | β Cnb1KO mice develop age-dependent diabetes mellitus.

a, Random-fed blood glucose levels in ageing β Cnb1KO (filled symbols) and littermate control (open symbols) mice. **b**, Glucose tolerance tests were performed on 12-week-old mice; symbols as in **a**. **c**, Serum insulin levels in random-fed 12-week-old β Cnb1KO (filled bars) and control (open bars) mice. **d**, **e**, Total pancreatic insulin (**d**) and glucagon (**e**) content in 12-week-old β Cnb1KO (filled bars) and control (open bars) mice. **f**, Body mass in ageing β Cnb1KO (filled symbols) and littermate control (open symbols) mice. **g**, β -cell mass in ageing β Cnb1KO (filled bars) and littermate control (open bars) mice. $n = 3$ –6 mice per genotype. **h**, BrdU incorporation by proliferating β -cells in β Cnb1KO (filled bars) and littermate control (open bars) mice. $n = 3$ –6 mice per genotype. All data are from both male and female mice and are presented as means $\pm$ s.e.m. $n = 10$ –20 animals per genotype unless otherwise noted. Asterisk, $P < 0.05$; two asterisks, $P < 0.01$; three asterisks, $P < 0.001$.

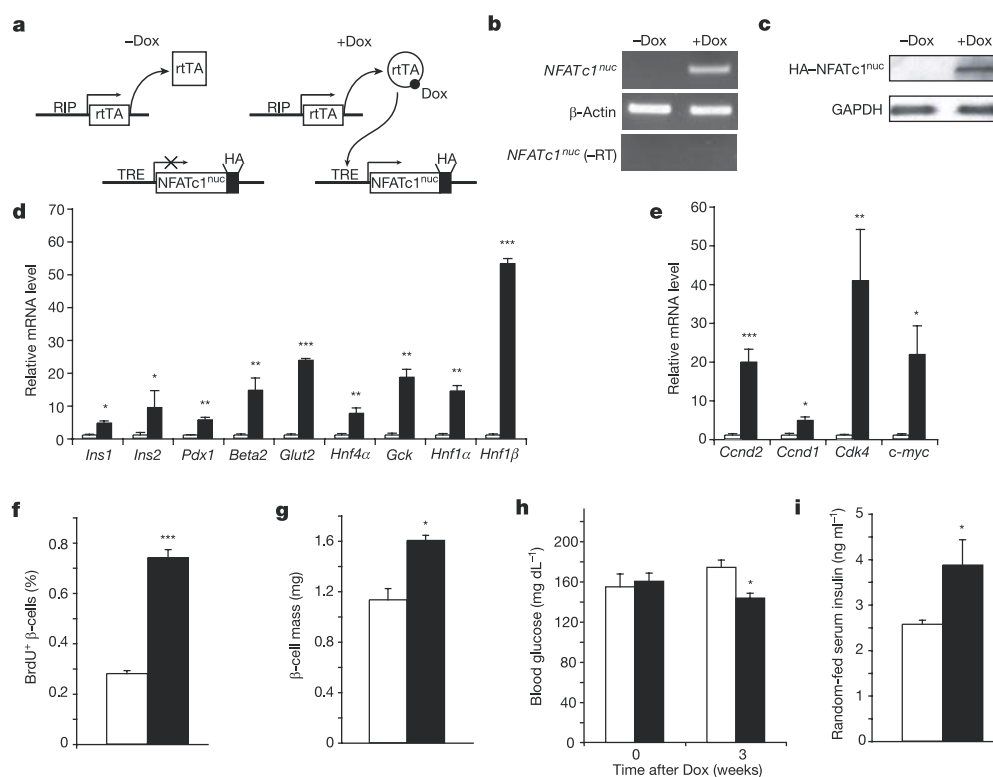
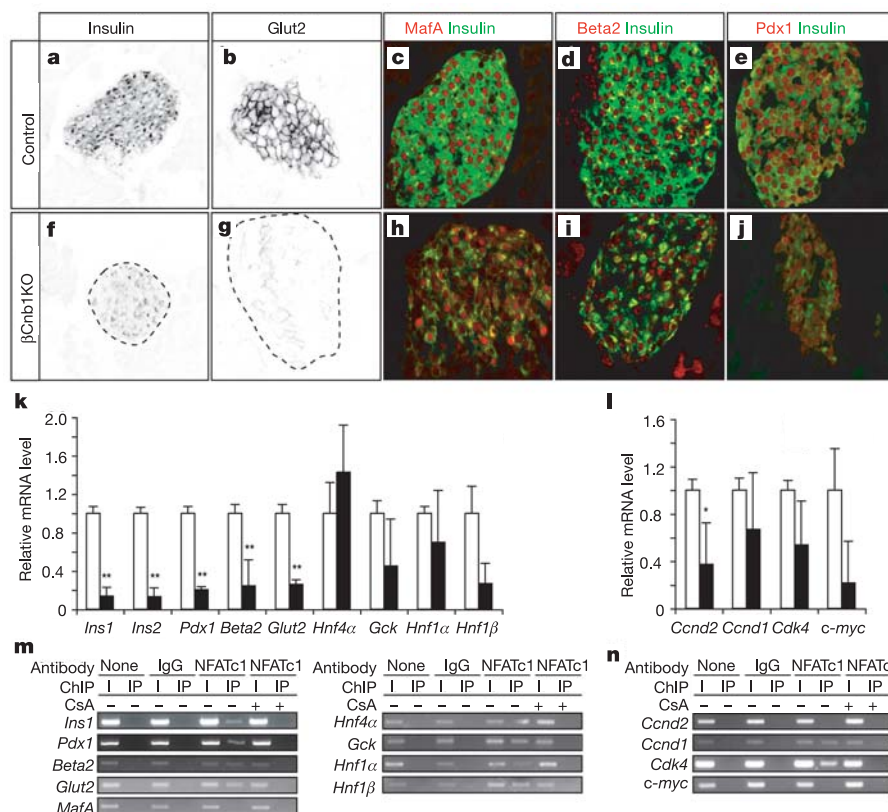


Figure 4 | Active NFATc1 induces the expression of essential β -cell markers, increases β -cell proliferation and mass, and induces hyperinsulinaemia. **a**, Diagram of Tet-On genetic system. **b**, RT-PCR demonstrating *NFATc1*^{nuc} induction in Dox-treated islets from RIP-rtTA; *NFATc1*^{nuc} mice. **c**, Western blot analysis demonstrating the detection of haemagglutinin (HA)-tagged NFATc1^{nuc} with anti-HA antisera after Dox induction. GAPDH, glyceraldehyde-3-phosphate dehydrogenase. **d, e**, Real-time RT-PCR quantification of characteristic β -cell gene products (**d**) or cell-cycle regulators (**e**) from RIP-rtTA; *NFATc1*^{nuc} islets treated (filled bars) or not

(open bars) with Dox. **f, g**, Islet β -cell BrdU incorporation (**f**) and β -cell mass (**g**) measured after treatment with (filled bar) or without (open bar) continuous exposure to Dox for 3 weeks in RIP-rtTA; *NFATc1*^{nuc} mice. **h**, Blood glucose levels in RIP-rtTA; *NFATc1*^{nuc} mice before and 3 weeks after *NFATc1*^{nuc} induction with Dox (filled bars) or in control RIP-rtTA; *NFATc1*^{nuc} mice not exposed to Dox (open bars). **i**, Serum insulin levels in random-fed RIP-rtTA; *NFATc1*^{nuc} mice with (filled bar) or without (open bar) continuous exposure to Dox for 3 weeks. $n = 5–7$ mice per group. Data are presented as means $\pm$ s.e.m. Asterisk, $P < 0.05$; two asterisks, $P < 0.01$; three asterisks, $P < 0.001$.

regulate the expression of essential β -cell factors *in vivo*. To test this hypothesis further, we generated mice that conditionally expressed activated NFATc1 in β -cells. We used an NFATc1 variant containing 21 amino-terminal serine-to-alanine substitutions, which localizes to the nucleus in a calcineurin-independent manner (NFATc1^{nuc}; refs 22, 23). We generated mice that express the reverse tetracycline transactivator (rtTA) from RIP²⁴ and NFATc1^{nuc} from the tetracycline-responsive element (TRE) promoter. Administration of doxycycline (Dox) to RIP-rtTA; NFATc1^{nuc} mice allows rtTA binding to the TRE promoter, which stimulates NFATc1^{nuc} expression (Fig. 4a). In islets isolated from 8-week-old RIP-rtTA; NFATc1^{nuc} mice and cultured *in vitro*, NFATc1^{nuc} mRNA and protein were observed only after the addition of Dox to the culture medium (Fig. 4b, c), demonstrating inducible expression of NFATc1^{nuc}. RIP-rtTA; NFATc1^{nuc} islets exposed to Dox for 48 h had increased expression of *Ins1*, *Ins2*, *Hnf4 α* , *Gck*, *Hnf1 α* , *Pdx1*, *Hnf1 β* , *Beta2*, *Glut2* and *MafA* (Fig. 4d). Expression of mRNA encoding the cell-cycle regulators *Cdk4*, *Ccnd1*, *Ccnd2* and *c-myc* was also increased on the induction of NFATc1^{nuc} (Fig. 4e). In contrast, expression of *GAPDH* and *p21* was not detectably altered (Supplementary Fig. 9). Furthermore, exposure of islets from RIP-rtTA or NFATc1^{nuc} single transgenic mice to Dox did not induce changes in gene expression (Supplementary Fig. 10). Together, these data show that active NFATc1 is sufficient to increase the expression of genes that regulate β -cell proliferation and function.

To assess NFATc1^{nuc} effects *in vivo*, we measured β -cell proliferation and mass in 8–10-week-old RIP-rtTA; NFATc1^{nuc} mice. *In vivo* islet NFATc1^{nuc} induction in Dox-treated mice was confirmed by RT-PCR and western blot analysis (data not shown). In contrast with control mice not exposed to Dox, Dox-treated RIP-rtTA; NFATc1^{nuc} mice had a 2.5-fold increase in numbers of BrdU-positive β -cells (Fig. 4f), an increase that corresponded with a 40% increase in β -cell mass (Fig. 4g). After 3 weeks of treatment with Dox, RIP-rtTA; NFATc1^{nuc} mice had a 20% reduction in random-fed blood glucose levels in comparison with RIP-rtTA; NFATc1^{nuc} controls never exposed to Dox (Fig. 4h). Consistent with this hypoglycaemia was our observation of increased serum insulin levels in Dox-treated RIP-rtTA; NFATc1^{nuc} mice in comparison with controls (Fig. 4i). Thus, conditional activation of NFATc1 in β -cells stimulated proliferation and increased insulin secretion. Moreover, RIP-rtTA; NFATc1^{nuc} mice are a unique strain that permits the conditional expansion of functional adult β -cells.

We postulated that impaired NFAT activity might promote diabetes pathogenesis in β Cnb1KO mice; we therefore tested whether NFATc1^{nuc} expression in β -cells of β Cnb1KO mice could prevent β -cell failure and diabetes pathogenesis. We intercrossed mice to generate RIP-Cre; Cnb1^{ff}; RIP-rtTA; TRE-NFATc1^{nuc} (abbreviated as β Cnb1KO; c1^{nuc}) and littermate control mice. Cnb1 mRNA levels were decreased to similar degrees in islets from both β Cnb1KO and β Cnb1KO; c1^{nuc} mice (Fig. 5a), indicating efficient Cnb1^{ff} inactivation in β Cnb1KO; c1^{nuc} mice. Treatment of β Cnb1KO; c1^{nuc} mice with Dox from 6 weeks of age resulted in detectable islet NFATc1^{nuc} expression, in contrast to controls, including β Cnb1KO; TRE-NFATc1^{nuc} littermates that lacked the RIP-rtTA transgene (Fig. 5b). Unlike β Cnb1KO littermates, which developed diabetes by 10 weeks, β Cnb1KO; c1^{nuc} mice maintained normal glucose levels during random feeding (Fig. 5c). Similarly, fasted β Cnb1KO; c1^{nuc} mice challenged with intraperitoneal glucose showed no evidence of impaired tolerance (Fig. 5d). Consistent with these findings was our observation that serum insulin levels in random-fed β Cnb1KO; c1^{nuc} mice did not differ significantly from those in non-diabetic control littermates (Fig. 5e). Thus, NFATc1^{nuc} expression in β -cells lacking Cnb1 was sufficient to maintain normoglycaemia. Moreover, islet and β -cell number, and β -cell proliferation, in β Cnb1KO; c1^{nuc} mice and littermate controls were indistinguishable (Fig. 5f, g). In comparison with islets from control mice, we found no significant decrease in mRNAs encoding β -cell growth regulators such as

insulin, Pdx1, Beta2 and cyclin D2 in β Cnb1KO; c1^{nuc} islets, in contrast to the severe decreases in these factors in islets from littermate β Cnb1KO mice (Fig. 5h). Taken together, these data indicate that NFATc activation promotes proliferation and essential endocrine activities in adult β -cells. Additional studies are needed

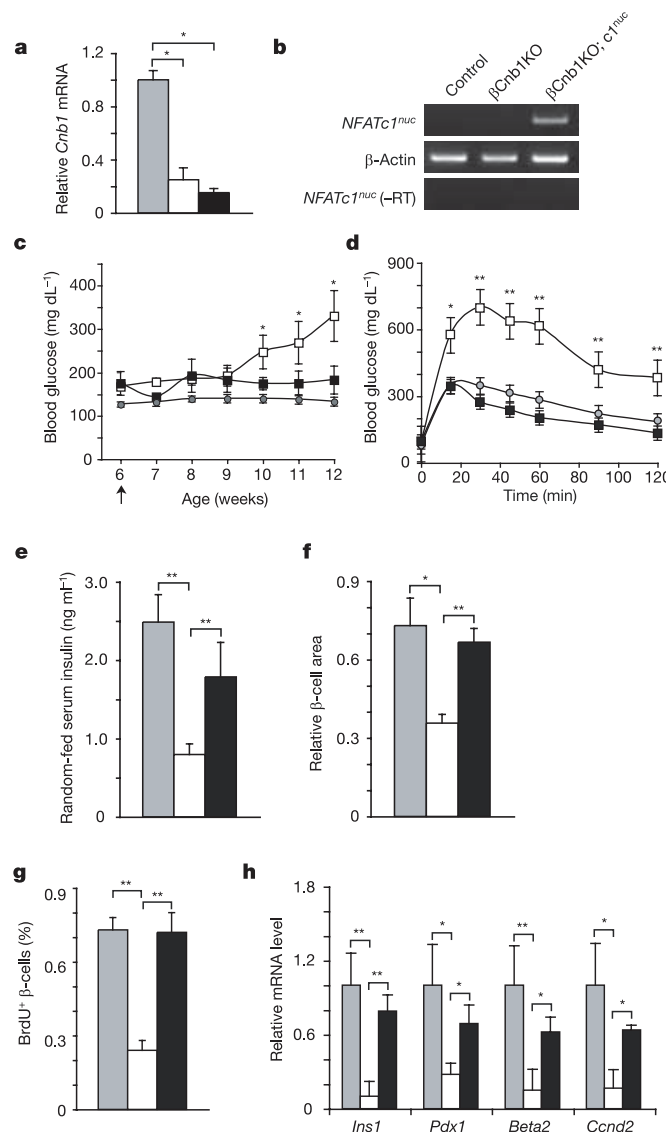


Figure 5 | Expression of active NFATc1 in Cnb1-deficient β -cells restores β -cell proliferation, mass and function, and prevents diabetes mellitus.

a, Real-time RT-PCR for *Cnb1* using handpicked islets from 12-week-old β Cnb1KO (white bars), β Cnb1KO; c1^{nuc} (black bars) and littermate control (grey bars) mice. **b**, Detection of NFATc1^{nuc} expression after administration of Dox in β Cnb1KO; c1^{nuc} and littermate control mice. **c**, Random-fed blood glucose levels in ageing β Cnb1KO (open squares), β Cnb1KO; c1^{nuc} (filled squares) and littermate control (grey circles) mice. Dox was started at the time indicated by the arrow. **d**, Glucose tolerance tests performed on 12-week-old β Cnb1KO (open squares), β Cnb1KO; c1^{nuc} (filled squares) and littermate control (grey circles) mice. **e**, Serum insulin levels in random-fed 12-week-old β Cnb1KO (white bars), β Cnb1KO; c1^{nuc} (black bars) and littermate control (grey bars) mice. **f**, β -cell mass in ageing β Cnb1KO (white bars), β Cnb1KO; c1^{nuc} (black bars) and littermate control (grey bars) mice. **g**, BrdU incorporation by proliferating β -cells in β Cnb1KO (white bars), β Cnb1KO; c1^{nuc} (black bars) and littermate control (grey bars) mice. **h**, Real-time RT-PCR for markers controlling β -cell proliferation with islets from 12-week-old β Cnb1KO (white bars), β Cnb1KO; c1^{nuc} (black bars) and littermate control (grey bars) mice. All data are from both male and female mice and are presented as means $\pm$ s.e.m. $n = 4$ –8 animals per genotype. Asterisk, $P < 0.05$; two asterisks, $P < 0.01$.

to identify β -cell roles for specific NFATc proteins, which have functional redundancy^{9,10}, and to address possible roles *in vivo* of other calcineurin substrates, such as Torc2 (ref. 25), in controlling postnatal β -cell expansion and function.

These studies reveal that calcineurin/NFAT signalling is critical for facultative β -cell proliferation and function *in vivo* (Supplementary Fig. 11). We speculate that this pathway may govern β -cell adaptation in other settings of increased insulin demand, including genetic insulin resistance, obesity and pregnancy, and during β -cell injury and regeneration. Our finding that this pathway regulates the expression of all six MODY factors indicates that dysregulated calcineurin/NFAT signalling could underlie some inherited forms of type 2 diabetes and pancreatic malformation²⁶. Moreover, our finding that disruption of calcineurin/NFAT function in β -cells is sufficient to induce overt diabetes provides unique evidence *in vivo* that iatrogenic diabetes in patients receiving calcineurin inhibitors³ stems from β -cell failure. The striking finding that activated NFATc1 is sufficient to increase β -cell proliferation and mass, leading to hyperinsulinaemia, raises the possibility that pathological activation of calcineurin/NFAT signalling may promote the pathogenesis of disordered β -cell growth, such as insulinomas or nesidioblastosis. If so, we speculate that calcineurin inhibitors might prove useful for treating the pathological overgrowth of β -cells. Our finding that small changes in nuclear occupancy by NFATc1 can markedly increase β -cell proliferation and insulin secretion indicates that pharmacological or genetic activation of NFATc proteins might be useful for reversing β -cell failure, promoting β -cell differentiation from progenitor cells, or improving the function of transplantable islets.

METHODS

Animals and physiological studies. Mouse maintenance, breeding and genotyping were as described^{8–10,22,24}. Blood glucose measurements, glucose tolerance tests, insulin secretion studies and insulin tolerance tests were performed as described^{27–29}. All measurements in this study are presented as means $\pm$ s.e.m. Significance was determined with Student's *t*-test, and *P* < 0.05 was considered significant.

Insulin levels were determined with an Ultrasensitive Insulin ELISA kit (Alpco). To measure pancreatic insulin and glucagon content, the pancreas was isolated, homogenized in acid alcohol, and extracted overnight at 4 °C. Insulin content was determined by ELISA as above, and glucagon levels were measured by radioimmunoassay (Linco Research Immunoassay; Linco Research). For all Tet-On experiments, Dox (1 g l⁻¹) was added to the drinking water and changed three times per week. Detailed methods are given in Supplementary Methods.

Islet isolation, gene expression studies, morphometry, immunohistology and chromatin methods. Islets were isolated, cultured and analysed as described²⁹. For Tet-On experiments, Dox (1 μ g ml⁻¹) was added to the growth medium to induce NFATc1^{mtc} expression. ChIP studies were performed as described²⁹, and DNA was detected with primer sequences listed in Supplementary Table 1. Detailed protocols are also given in Supplementary Methods.

Received 3 May; accepted 18 July 2006.

- Heit, J. J., Karnik, S. K. & Kim, S. K. Intrinsic regulators of pancreatic β -cell proliferation. *Annu. Rev. Cell Dev. Biol.* **22**, 311–338 (2006).
- Cozar-Castellano, I. et al. Molecular control of cell cycle progression in the pancreatic β -cell. *Endocr. Rev.* **27**, 356–370 (2006).
- Weir, M. R. & Fink, J. C. Risk for posttransplant diabetes mellitus with current immunosuppressive medications. *Am. J. Kidney Dis.* **34**, 1–13 (1999).
- Georgia, S. & Bhushan, A. β cell replication is the primary mechanism for maintaining postnatal β cell mass. *J. Clin. Invest.* **114**, 963–968 (2004).
- Kushner, J. A. et al. Cyclins D2 and D1 are essential for postnatal pancreatic β -cell growth. *Mol. Cell. Biol.* **25**, 3752–3762 (2005).
- Graef, I. A., Chen, F., Chen, L., Kuo, A. & Crabtree, G. R. Signals transduced by Ca²⁺/calcineurin and NFATc3/c4 pattern the developing vasculature. *Cell* **105**, 863–875 (2001).
- Crabtree, G. R. & Olson, E. N. NFAT signaling: choreographing the social lives of cells. *Cell* **109** (Suppl.), S67–S79 (2002).
- Neilson, J. R., Winslow, M. M., Hur, E. M. & Crabtree, G. R. Calcineurin B1 is essential for positive but not negative selection during thymocyte development. *Immunity* **20**, 255–266 (2004).
- Winslow, M. M., Gallo, E. M., Neilson, J. R. & Crabtree, G. R. The calcineurin

phosphatase complex modulates immunogenic B cell responses. *Immunity* **24**, 141–152 (2006).

- Herrera, P. L. Adult insulin- and glucagon-producing cells differentiate from two independent cell lineages. *Development* **127**, 2317–2322 (2000).
- Lawrence, M. C., Bhatt, H. S., Watterson, J. M. & Easom, R. A. Regulation of insulin gene transcription by a Ca²⁺-responsive pathway involving calcineurin and nuclear factor of activated T cells. *Mol. Endocrinol.* **15**, 1758–1767 (2001).
- Lawrence, M. C., Bhatt, H. S. & Easom, R. A. NFAT regulates insulin gene promoter activity in response to synergistic pathways induced by glucose and glucagon-like peptide-1. *Diabetes* **51**, 691–698 (2002).
- Lawrence, M. C., McGlynn, K., Park, B. H. & Cobb, M. H. ERK1/2-dependent activation of transcription factors required for acute and chronic effects of glucose on the insulin gene promoter. *J. Biol. Chem.* **280**, 26751–26759 (2005).
- Redmon, J. B., Olson, L. K., Armstrong, M. B., Greene, M. J. & Robertson, R. P. Effects of tacrolimus (FK506) on human insulin gene expression, insulin mRNA levels, and insulin secretion in HIT-T15 cells. *J. Clin. Invest.* **98**, 2786–2793 (1996).
- Thorens, B., Guillaum, M. T., Beermann, F., Burcelin, R. & Jaquet, M. Transgenic reexpression of GLUT1 or GLUT2 in pancreatic β cells rescues GLUT2-null mice from early death and restores normal glucose-stimulated insulin secretion. *J. Biol. Chem.* **275**, 23751–23758 (2000).
- O'Rahilly, S., Barroso, I. & Wareham, N. J. Genetic factors in type 2 diabetes: the end of the beginning? *Science* **307**, 370–373 (2005).
- Rane, S. G. et al. Loss of Cdk4 expression causes insulin-deficient diabetes and Cdk4 activation results in β -islet cell hyperplasia. *Nature Genet.* **22**, 44–52 (1999).
- Laybutt, D. R. et al. Overexpression of c-Myc in β -cells of transgenic mice causes proliferation and apoptosis, downregulation of insulin gene expression, and diabetes. *Diabetes* **51**, 1793–1804 (2002).
- Pelengaris, S., Khan, M. & Evan, G. I. Suppression of Myc-induced apoptosis in β cells exposes multiple oncogenic properties of Myc and triggers carcinogenic progression. *Cell* **109**, 321–334 (2002).
- Miyazaki, J. et al. Establishment of a pancreatic beta cell line that retains glucose-inducible insulin secretion: special reference to expression of glucose transporter isoforms. *Endocrinology* **127**, 126–132 (1990).
- Baksh, S. et al. NFATc2-mediated repression of cyclin-dependent kinase 4 expression. *Mol. Cell* **10**, 1071–1081 (2002).
- Beals, C. R., Clipstone, N. A., Ho, S. N. & Crabtree, G. R. Nuclear localization of NF-ATc by a calcineurin-dependent, cyclosporin-sensitive intramolecular interaction. *Genes Dev.* **11**, 824–834 (1997).
- Winslow, M. M. et al. Calcineurin/NFAT signaling in osteoblasts regulates bone mass. *Dev. Cell* **10**, 771–782 (2006).
- Smart, N. G. et al. Conditional expression of Smad7 in pancreatic β cells disrupts TGF- β signaling and induces reversible diabetes mellitus. *PLoS Biol.* **4**, e39 (2006).
- Screaton, R. A. et al. The CREB coactivator TORC2 functions as a calcium- and cAMP-sensitive coincidence detector. *Cell* **119**, 61–74 (2004).
- Aron, J. R. et al. NFAT dysregulation by increased dosage of DSCR1 and DYRK1A on chromosome 21. *Nature* **441**, 595–600 (2006).
- Bruning, J. C. et al. Development of a novel polygenic model of NIDDM in mice heterozygous for *IR* and *IRS-1* null alleles. *Cell* **88**, 561–572 (1997).
- Kulkarni, R. N. et al. PDX-1 haploinsufficiency limits the compensatory islet hyperplasia that occurs in response to insulin resistance. *J. Clin. Invest.* **114**, 828–836 (2004).
- Karnik, S. K. et al. Menin regulates pancreatic islet growth by promoting histone methylation and expression of genes encoding p27Kip1 and p18INK4c. *Proc. Natl Acad. Sci. USA* **102**, 14659–14664 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. McLean, M. McLane and J. Morillo for technical assistance; R. Kulkarni, A. Xu and G. Barsh for advice; and members of the Crabtree and Kim laboratories for discussions and critical readings of this manuscript. J.J.H. is a student in the Stanford Medical Scientist Training Program and is additionally supported by an American Diabetes Association (ADA) Medical Scholars grant. M.M.W. is supported by a Stanford graduate fellowship and a Howard Hughes Medical Institute (HHMI) Predoctoral Fellowship. This work was supported by awards to G.R.C. from HHMI and the NIH, and to S.K.K. from the NIH, the Biomedical Scholars Program of the Pew Charitable Trusts, and the Stephen and Caroline Kaufer Fund for Neuroendocrine Tumor Research.

Author Contributions J.J.H., Å.A.A. and S.K.K. generated the hypotheses and designed the experiments. J.R.N., M.M.W. and G.R.C. generated and helped validate *Cnb1*^f, *Cnb1*^d and TRE-NFATc1^{mtc} transgenic mice. X.G., Å.A.A. and J.J.H. performed experiments. J.J.H. and S.K.K. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.K.K. (seungkim@stanford.edu).

LETTERS

PD-1 expression on HIV-specific T cells is associated with T-cell exhaustion and disease progression

Cheryl L. Day^{1,2,3*}, Daniel E. Kaufmann^{2*}, Photini Kiepiela¹, Julia A. Brown⁴, Eshia S. Moodley¹, Sharon Reddy¹, Elizabeth W. Mackey², Joseph D. Miller⁵, Alasdair J. Leslie³, Chantal DePierres¹, Zenele Mncube¹, Jaikumar Duraiswamy⁵, Baogong Zhu⁴, Quentin Eichbaum², Marcus Altfeld², E. John Wherry⁶, Hoosen M. Coovadia¹, Philip J. R. Goulder^{1,2,3}, Paul Klenerman³, Rafi Ahmed⁵, Gordon J. Freeman⁴ & Bruce D. Walker^{1,2,7}

Functional impairment of T cells is characteristic of many chronic mouse and human viral infections. The inhibitory receptor programmed death 1 (PD-1; also known as PDCD1), a negative regulator of activated T cells^{1–4}, is markedly upregulated on the surface of exhausted virus-specific CD8 T cells in mice⁵. Blockade of this pathway using antibodies against the PD ligand 1 (PD-L1, also known as CD274) restores CD8 T-cell function and reduces viral load⁵. To investigate the role of PD-1 in a chronic human viral infection, we examined PD-1 expression on human immunodeficiency virus (HIV)-specific CD8 T cells in 71 clade-C-infected people who were naive to anti-HIV treatments, using ten major histocompatibility complex (MHC) class I tetramers specific for frequently targeted epitopes. Here we report that PD-1 is significantly upregulated on these cells, and expression correlates with impaired HIV-specific CD8 T-cell function as well as predictors of disease progression: positively with plasma viral load and inversely with CD4 T-cell count. PD-1 expression on CD4 T cells likewise showed a positive correlation with viral load and an inverse correlation with CD4 T-cell count, and blockade of the pathway augmented HIV-specific CD4 and CD8 T-cell function. These data indicate that the immunoregulatory PD-1/PD-L1 pathway is operative during a persistent viral infection in humans, and define a reversible defect in HIV-specific T-cell function. Moreover, this pathway of reversible T-cell impairment provides a potential target for enhancing the function of exhausted T cells in chronic HIV infection.

Recent evidence from experimental lymphocytic choriomeningitis virus (LCMV) infection indicates a crucial role of the PD-1/PD-L1 pathway in inhibiting the function of virus-specific CD8 T cells in chronic viral infections in mice⁵. To address the potential role of this pathway in a chronic human viral infection associated with persistent viraemia and impaired T-cell function^{6–13}, we examined peripheral blood from a cohort¹⁴ of over 500 people with chronic untreated HIV infection in KwaZulu Natal Province in South Africa, where seroprevalence rates are in excess of 30% in certain age groups.

A panel of ten MHC class I tetramers—based on prevalent human leukocyte antigen (HLA) alleles and frequently targeted HIV-1 clade C epitopes¹⁴—was synthesized (Supplementary Table 1), allowing the

direct visualization of surface PD-1 expression on tetramer-positive (tetramer⁺) cells (Fig. 1a). A subset of 71 people naive for anti-retroviral therapy was studied on the basis of expression of relevant HLA alleles, and a total of 126 individual tetramer responses were examined (Fig. 1b). PD-1 expression was readily apparent on these tetramer⁺ cells, and was significantly higher than in the total CD8 T-cell population ($P < 0.0001$). In turn, PD-1 expression on both tetramer⁺ CD8 T cells and the total CD8 T-cell population was significantly higher than in HIV-seronegative controls. Significantly less PD-1 was expressed on cytomegalovirus (CMV)-specific cells compared with HIV-specific CD8 T cells ($P = 0.0002$), whereas PD-1 expression on cells specific for a single lytic Epstein–Barr virus (EBV) epitope tested was also at high levels.

PD-1 expression was also analysed on CMV-specific, EBV-specific and vaccinia-virus-specific CD8 T cells from HIV-seronegative controls, and found to be intermediate on CMV-specific cells, high for a lytic EBV epitope, and low on vaccinia-virus-specific CD8 T cells in 18 healthy individuals (Fig. 1c), indicating a relationship between ongoing antigen exposure and PD-1 expression. High PD-1 expression on HIV-specific CD8 T cells was unique as we did not see upregulation of the inhibitory receptor CTLA-4 (cytotoxic T-lymphocyte-associated protein 4; data not shown), consistent with previous reports¹⁵.

Next, we determined whether there was evidence for epitope-specific differences in PD-1 expression. Each HIV tetramer was used to stain peripheral blood mononuclear cells (PBMCs) from 3 to 29 individuals of the appropriate HLA type, revealing a range of PD-1 expression levels on tetramer⁺ cells (Fig. 1d); for a given epitope, the median percentage of PD-1-expressing tetramer⁺ cells ranged from 68% to 94% (Fig. 1e). Of the 71 people examined, 16 individuals had multiple tetramer responses—13 of whom showed different patterns of PD-1 expression depending on the epitope (Fig. 1f). These data indicate that PD-1 may be differentially expressed on contemporaneous epitope-specific CD8 T cells from a single person, perhaps consistent with data indicating epitope-specific differences in antiviral efficacy^{16–18}.

Because HIV-specific CD8 T cells also show impaired proliferative capacity^{6,7}, we stimulated carboxyfluorescein diacetate succinimidyl

¹HIV Pathogenesis Programme, Doris Duke Medical Research Institute, University of KwaZulu Natal, Durban 4013, South Africa. ²Partners AIDS Research Center, Massachusetts General Hospital and Division of AIDS, Harvard Medical School, Boston, Massachusetts 02115, USA. ³Nuffield Department of Medicine, The Peter Medawar Building for Pathogen Research, Oxford University, Oxford OX1 35Y, UK. ⁴Department of Medical Oncology, Dana-Farber Cancer Institute, Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁵Emory Vaccine Center and Department of Microbiology & Immunology, Emory University School of Medicine, Atlanta, Georgia 30322, USA. ⁶Immunology Program, The Wistar Institute, Philadelphia, Pennsylvania 19104, USA. ⁷Howard Hughes Medical Institute, Chevy Chase, Maryland 20185, USA.

*These authors contributed equally to this work.

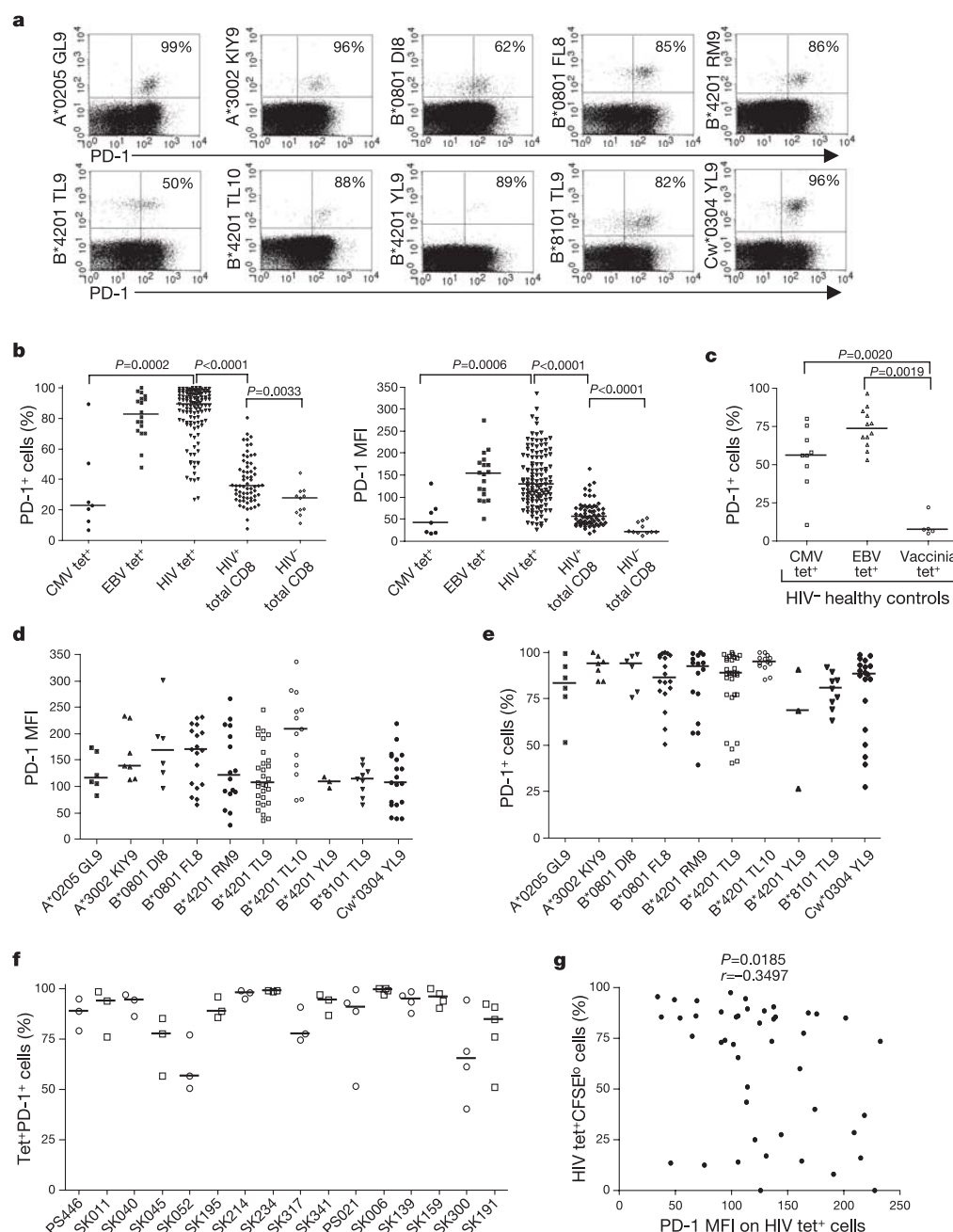


Figure 1 | PD-1 is upregulated on HIV-specific CD8 T cells.

a, Representative PD-1 expression using ten HIV tetramers specific for epitopes frequently targeted in chronic clade C HIV infection. Values in the upper right quadrant indicate the percentage of tetramer⁺ cells that express PD-1. **b**, Percentage and mean fluorescence intensity (MFI, relative fluorescence units) of PD-1 expression on HIV-specific CD8 T cells compared with the total CD8 T-cell population in individuals naive for anti-retroviral therapy ($n = 71$), and on the total CD8 T-cell population in HIV-infected subjects ($n = 71$) versus HIV-seronegative controls ($n = 11$). Horizontal bars indicate the median percentage of PD-1-positive cells. Statistical comparisons were made using the Mann-Whitney test. **c**, Percentage PD-1 expression on CMV-specific, EBV-specific and vaccinia-virus-specific CD8 T cells from HIV-seronegative controls ($n = 18$). Horizontal bars indicate the median percentage of PD-1-expressing tetramer⁺ cells. Statistical comparisons were made using the Mann-Whitney test. **d**, PD-1 expression on tetramer⁺ cells by epitope specificity. Horizontal bars indicate the median PD-1 MFI values. **e**, Median percentage of PD-1-expressing tetramer⁺ cells by epitope specificity. **f**, Variation in the percentage of PD-1-expressing cells in different epitope-specific populations within individuals with multiple detectable responses. Horizontal bars indicate the median percentage of PD-1-expressing HIV tetramer⁺ cells in each individual. **g**, Inverse relationship between PD-1 expression and proliferative capacity of HIV-specific CD8 T cells ($n = 45$; correlation statistics were analysed using the Spearman correlation).

ester (CFSE)-labelled PBMC with HIV peptides for 6 days to analyse HIV-specific CD8 T-cell proliferation in relation to PD-1 expression on HIV tetramer⁺ cells (Supplementary Fig. 1). We found a significant inverse correlation between the percentage of proliferating cells (tetramer⁺CFSE¹⁰) and the level of expression of PD-1 on HIV tetramer⁺ cells in the freshly isolated PBMCs ($P = 0.0185$; Fig. 1g). These data indicate that HIV-specific CD8 T cells expressing high amounts of PD-1 are more functionally exhausted in that they are less able to respond to cognate antigen *in vitro*.

Next, we assessed the relationship between PD-1 expression, plasma viral load, and CD4 T-cell counts—the latter two being predictors of HIV disease progression. There was no correlation between tetramer⁺ CD8 T cells and viral load or CD4 T-cell count (Fig. 2a, b, right panels). In contrast, we found significant positive correlations between viral load and both the mean fluorescence intensity (MFI; $P < 0.0001$; Fig. 2a, left panel) and percentage PD-1 expression ($P = 0.0001$; Supplementary Fig. 2) on HIV tetramer⁺ cells, and inverse correlations between CD4 T-cell count

and both the MFI ($P = 0.0103$; Fig. 2b, left panel) and the percentage of PD-1-expressing HIV tetramer⁺ cells ($P = 0.0031$; Supplementary Fig. 2). Because the tetramers tested probably represent only a fraction of the HIV-specific CD8 T-cell population in these subjects, we also examined PD-1 expression on total CD8 T cells, and found significant positive correlations between viral load and PD-1 expression ($P < 0.0001$ by PD-1 MFI (Fig. 2c); $P = 0.0002$ by the percentage of PD-1-expressing cells (Supplementary Fig. 2)), and inverse correlations between CD4 T-cell count and PD-1 expression ($P = 0.0002$ by PD-1 MFI (Fig. 2d); $P = 0.0042$ by the percentage of PD-1-expressing cells (Supplementary Fig. 2)). These data indicate that increasing amounts of antigen in chronic HIV infection are associated with increased expression of PD-1 on CD8 T cells. Moreover, they provide the first clear association, in a large study including analysis of multiple epitopes, between a marker on HIV-specific CD8 T cells and both viral load and CD4 T-cell count.

We also examined PD-1 expression in relation to manipulation of antigen exposure by treatment with antiviral therapy. Before

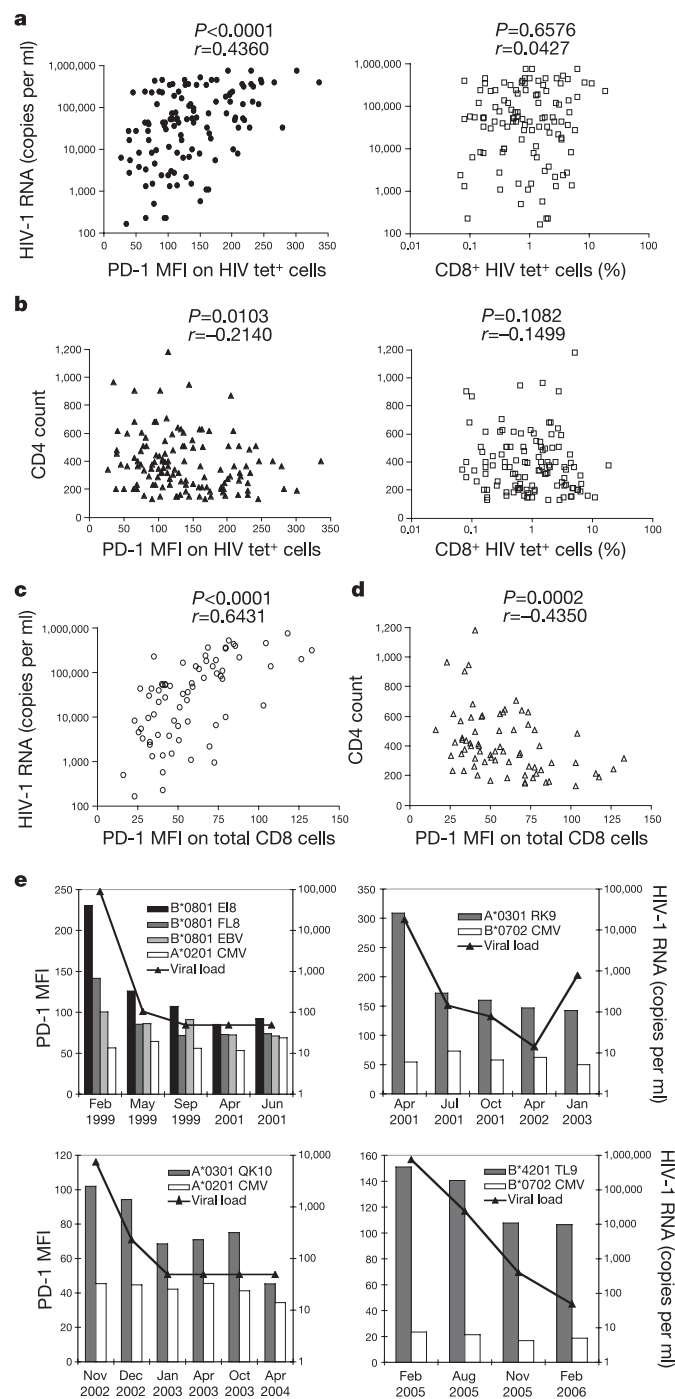


Figure 2 | PD-1 expression is associated with HIV disease progression.

a, There is no correlation between the number of HIV-specific CD8 T cells (as measured by tetramer staining) and plasma viral load (right panel), whereas there is a positive correlation between the MFI of PD-1 expression on tetramer⁺ cells and plasma viral load (left panel). **b**, There is no correlation between the frequency of HIV tetramer⁺ cells and CD4 T-cell count (right panel), whereas there is an inverse correlation between the MFI of PD-1 expression on HIV tetramer⁺ cells and CD4 T-cell count (left panel). **c**, The MFI of PD-1 expression on the total CD8 T-cell population correlates positively with plasma viral load. **d**, The MFI of PD-1 expression on the total CD8 T-cell population is inversely correlated with CD4 T-cell count. Correlation statistics for **a–d** were analysed using the Spearman correlation. **e**, Following initiation of anti-retroviral therapy, PD-1 expression decreases on HIV-specific, but not EBV- or CMV-specific, CD8 T cells. MFI of PD-1 expression on tetramer⁺ cells (grey and white bars) is indicated on the left y-axes, and the number of HIV-1 RNA copies per ml of plasma (black line) is indicated on the right y-axes.

initiation of therapy, high levels of PD-1 expression were detected on HIV-specific CD8 T cells in all four subjects analysed. Anti-retroviral treatment resulted in the dramatic decline of detectable plasma viral load, coincident with a decrease in the MFI of PD-1 expression on HIV tetramer⁺ cells (Fig. 2e). No decrease on either EBV- or CMV-specific CD8 T cells was observed in the same individuals. These data indicate that high levels of antigen in chronic HIV infection may drive continuous high-level expression of PD-1 on HIV-specific CD8 T cells, and that lower, although still relatively elevated, levels of PD-1 persist despite prolonged suppression of HIV when compared with CMV-specific CD8 T cells in the same people.

PD-1 expression was analysed in relation to phenotypic markers associated with CD8 T-cell memory status and function (Supplementary Fig. 3). The HIV tetramer⁺ PD-1-expressing cells also express high levels of CD27 and granzyme B, but very low levels of CD28, CCR7 and intracellular Ki67, low levels of CD45RA and perforin, and intermediate levels of CD57, CD62L and CD127. These data indicate that HIV-specific PD-1-positive CD8 T cells have an effector/effector memory phenotype, and are consistent with previous reports of skewed maturation of HIV-specific CD8 T cells^{19–21}.

Previous studies in experimental LCMV infection showed that *in vivo* blockade of PD-1/PD-L1 interaction using a blocking antibody to PD-L1 results in enhanced clonal expansion, effector cytokine production, and cytotoxicity in exhausted LCMV-specific CD8 T cells resulting in a notable reduction in viral load in mice⁵. As we observed an inverse correlation between the proliferative capacity of HIV-specific CD8 T cells and the expression levels of PD-1, we examined whether blockade of the PD-1/PD-L1 pathway could enhance the ability of these cells to expand following stimulation with peptide *in vitro*. No effect of blockade was apparent in intracellular cytokine staining assays following a 6-h stimulation with peptide (data not shown). Therefore, we incubated freshly isolated PBMCs with medium alone or medium containing anti-PD-L1 antibody, and exposed the cells to HIV peptides for 6 days, an example of which is shown in Fig. 3a. In this case, stimulation with an HIV peptide (TL9) alone resulted in a 4.8-fold expansion of B*4201 TL9 tetramer⁺ cells, whereas stimulation with TL9 peptide in the presence of anti-PD-L1 antibody further enhanced proliferation of TL9-specific cells, resulting in a 10.3-fold increase in tetramer⁺ cells. Similar assays were performed on a total of 28 samples from 25 different subjects, and a significant increase in the expansion of HIV-specific CD8⁺ T cells was observed in the presence of peptide plus anti-PD-L1 antibody, as compared with expansion following stimulation with peptide alone (Fig. 3b; $P < 0.0001$). The fold-increase of tetramer⁺ cells in the presence of anti-PD-L1 antibody varied by individual, and by epitope within a given individual (Supplementary Fig. 4), again suggesting epitope specific differences in the degree of functional exhaustion of these responses.

Functionality, as measured by *in vitro* cytokine production by HIV-specific CD8 T cells following blockade of the PD-1 pathway, was also assessed. PBMCs from 14 individuals were incubated with isotype control antibodies, anti-PD-L1, anti-PD-L2, or anti-PD-L1 and anti-PD-L2 together, in the presence or absence of an HIV peptide targeted by CD8 T cells in that person. On day 6, the cells were re-stimulated for 12 h with the specific peptide and analysed for interferon (IFN)- γ production by intracellular cytokine staining. Representative data are shown in Fig. 3c. Incubation of PBMCs with the HIV peptide YL9 in the presence of either anti-PD-L1 or anti-PD-L2 resulted in a twofold increase in IFN- γ -producing cells, and incubation of both anti-PD-L1 and anti-PD-L2 together with peptide resulted in a 3.8-fold increase in the frequency of IFN- γ -producing CD8 T cells, as compared with peptide plus isotype control antibodies. For 13 out of 14 individuals tested, incubation of PBMCs with peptide and anti-PD-L1 and anti-PD-L2 antibodies together resulted in a significant increase in IFN- γ -producing CD8 T cells compared with PBMCs incubated with peptide and isotype control antibodies ($P = 0.0017$; Fig. 3d).

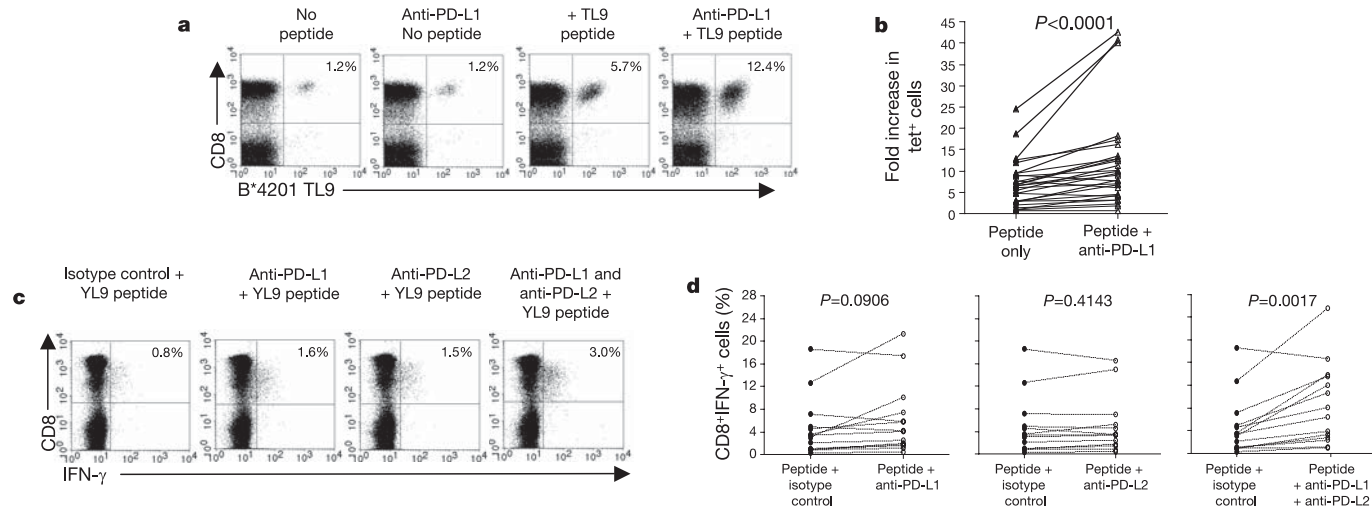


Figure 3 | Blockade of the PD-1/PD-L1 pathway significantly increases expansion of tetramer⁺ cells and frequency of HIV-specific, IFN- γ -producing CD8 T cells. **a**, Representative data of expansion of tetramer⁺ cells in the presence of anti-PD-L1 antibody. **b**, Summary data for expansion of tetramer⁺ cells in the presence of anti-PD-L1 blocking antibody ($n = 28$; statistical comparisons were made using the Wilcoxon matched pairs test). **c**, Representative intracellular cytokine staining data

from a Cw*0304-positive subject following a 6-day incubation with YL9 peptide and either isotype control antibodies, anti-PD-L1, anti-PD-L2, or anti-PD-L1 and anti-PD-L2 antibodies together. **d**, Dual blockade of the PD-1 pathway by both anti-PD-L1 and anti-PD-L2 results in a significant increase in the frequency of HIV-specific, IFN- γ -producing CD8 T cells ($n = 14$; statistical comparisons were made using the Wilcoxon matched pairs test).

Because of the link between CD4 T-cell function and control of viraemia in chronic viral infections (reviewed in ref. 22), we examined the expression of PD-1 on CD4 T cells in HIV-infected subjects. Similar to the findings with CD8 T cells, there was a significant inverse correlation between absolute CD4 T-cell count and PD-1 expression on CD4 T cells (Fig. 4a; $P < 0.0001$), and a positive correlation with viral load (Fig. 4b; $P = 0.0091$), which was apparent both for the MFI and percentage of PD-1-expressing CD4 cells (Fig. 4b and data not shown).

We also investigated the effect of blockade of the PD-1 pathway on HIV-specific CD4 T-cell function using anti-PD-L1 antibodies in a CFSE proliferation assay. Blockade of the PD-1 pathway resulted in an increase in HIV-specific proliferation, as measured by CFSE dye dilution (Fig. 4c). Moreover, in 6 out of 10 subjects there was no

detectable proliferation to HIV p24 protein in the absence of anti-PD-L1 antibody, but vigorous proliferation in the presence of anti-PD-L1 antibody in 5 out of these 6 subjects ($P = 0.0039$; Fig. 4d), indicating that these HIV-specific CD4 T cells are present but functionally impaired, and that function can be restored following blockade of the PD-1 pathway.

These data demonstrate that the PD-1 inhibitory pathway, in addition to regulating T-cell responses to self antigens and to viral antigens in mice, also regulates both CD4 and CD8 T-cell responses to a chronic human viral pathogen characterized by persistent viraemia and ultimately profound immune suppression. PD-1 expression was upregulated on total CD8 and CD4 T cells in people with chronic HIV infection naive for anti-retroviral therapy, and tetramer staining showed that this upregulation was markedly higher

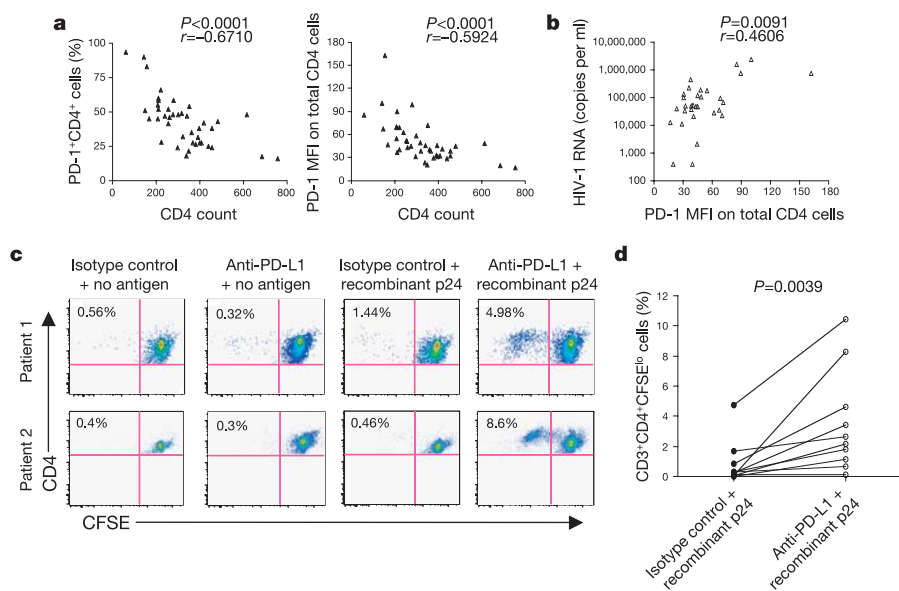


Figure 4 | Effect of PD-1 expression on CD4 T cells. **a**, The percentage and MFI of PD-1 expression on the total CD4 T-cell population in HIV-infected individuals inversely correlate with absolute CD4 T-cell counts ($n = 41$; correlation statistics were analysed using the Spearman correlation). **b**, The MFI of PD-1 expression on the total CD4 T-cell population correlates positively with HIV-1 viral load (correlation statistics were analysed using the Spearman correlation). **c**, Blockade of the PD-1 pathway significantly increases proliferation of HIV-specific CD4 T cells (top row, viral load = 635 HIV-1 RNA copies per ml plasma; CD4 T-cell count = 344), as well as restores proliferative capacity of HIV-specific CD4 T cells in a subject with no detectable responses following stimulation with recombinant p24 antigen alone (bottom row, viral load = 223,000 HIV-1 RNA copies per ml plasma; CD4 T-cell count = 446). Values in the upper left quadrant indicate the percentage of CD4⁺CD3⁺CFSE^{lo} cells. **d**, Summary data of proliferative responses to recombinant p24 antigen in the presence and absence of anti-PD-L1 antibody ($n = 10$; statistical comparisons were made using the Wilcoxon matched pairs test).

in the HIV-specific CD8 T-cell population. PD-1 expression on both CD4 and CD8 T cells correlated positively with viral load and negatively with CD4 T-cell count, providing the first clear evidence of a correlation between the adaptive T-cell response and disease progression parameters in this infected population, as well as a link between CD4 and CD8 T-cell function in the presence of ongoing viraemia.

Previous work has shown that PD-L1 is upregulated and B7-2 (also known as CD86) is downregulated in HIV infection, providing increased inhibitory and reduced activating signals for T cells²³. Together, these data indicate that increasing levels of antigenaemia in HIV infection results in immunoregulatory signals mediated through PD-1 and its ligands that impair T-cell function, providing a mechanism to explain the inability of HIV-specific cellular immune responses to effectively control viraemia. Whether blocking this pathway, which resulted in the augmentation of HIV-specific CD4 and CD8 T-cell function *in vitro*, would lead to enhanced control of viraemia (as observed in mice with LCMV infection⁵), or whether disruption of this regulatory T-cell pathway would have adverse effects, will require additional study.

METHODS

Study subjects. Peripheral blood was obtained from clade-C-virus-infected subjects in KwaZulu Natal, South Africa, all of whom were anti-retroviral therapy naive at the time of analysis. Median viral load was 42,400 HIV-1 RNA copies per ml plasma (range 163–750,000), and the median absolute CD4 T-cell count was 363 (range 129–1,179). An additional 41 clade-C-infected subjects were included for analysis of PD-1 expression on CD4 T cells. Ten subjects were recruited from Boston, USA, for functional analysis of PD-1 blockade in CD4 T cells. Longitudinal cryopreserved PBMC samples were analysed from four additional subjects who received highly active anti-retroviral therapy. All subjects gave written informed consent for the study, which was approved by local institutional review boards.

MHC class I tetramers. Ten HIV MHC class I tetramers were synthesized as described previously²⁴ (Supplementary Table 1).

CFSE proliferation assays. CFSE assays were performed as described previously⁷, using a final concentration of 0.2 µg ml⁻¹ peptide for CD8 T-cell responses, and a final concentration of 5 µg ml⁻¹ of recombinant p24 protein for CD4 T-cell responses.

PD-1/PD-L1 blockade. Freshly isolated PBMCs were incubated with either isotype control antibody (IgG2b clone MPC.11; 10 µg ml⁻¹), purified anti-PD-L1^{25,26} (10 µg ml⁻¹), isotype control antibody plus 0.2 µg ml⁻¹ peptide for CD8 T cells or 5 µg ml⁻¹ p24 protein for CD4 T cells, or anti-PD-L1 antibody plus antigen. Cells were incubated for 6 days, stained with surface antibodies and tetramers, and analysed by flow cytometry.

Analysis of cytokine production. Freshly isolated PBMCs were stimulated for 6 days with or without peptide in the presence of isotype control antibody, anti-PD-L1, anti-PD-L2, or anti-PD-L1 and anti-PD-L2 together. On day 6, cells were re-stimulated with peptide (1 µg ml⁻¹) overnight in the presence of brefeldin A. Cells were stained extracellularly with anti-CD8-APC (allophycocyanin) and ViaProbe (Becton Dickinson), and intracellularly with anti-IFN-γ-FITC (fluorescein isothiocyanate).

Statistical analysis. Spearman rank-correlation, Mann–Whitney and Wilcoxon matched pairs tests were performed using GraphPad Prism version 4.0a. All tests were two-tailed and *P*-values of *P* < 0.05 were considered significant.

Received 16 June; accepted 28 July 2006.

Published online 20 August 2006.

1. Ishida, Y., Agata, Y., Shibahara, K. & Honjo, T. Induced expression of PD-1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death. *EMBO J.* **11**, 3887–3895 (1992).
2. Nishimura, H., Nose, M., Hiai, H., Minato, N. & Honjo, T. Development of lupus-like autoimmune diseases by disruption of the PD-1 gene encoding an ITIM motif-carrying immunoreceptor. *Immunity* **11**, 141–151 (1999).
3. Sharpe, A. H. & Freeman, G. J. The B7–CD28 superfamily. *Nature Rev. Immunol.* **2**, 116–126 (2002).
4. Chen, L. Co-inhibitory molecules of the B7–CD28 family in the control of T-cell immunity. *Nature Rev. Immunol.* **4**, 336–347 (2004).

5. Barber, D. L. *et al.* Restoring function in exhausted CD8 T cells during chronic viral infection. *Nature* **439**, 682–687 (2006; published online 28 December 2005).
6. Migueles, S. A. *et al.* HIV-specific CD8⁺ T cell proliferation is coupled to perforin expression and is maintained in nonprogressors. *Nature Immunol.* **3**, 1061–1068 (2002).
7. Lichterfeld, M. *et al.* Loss of HIV-1-specific CD8⁺ T cell proliferation after acute HIV-1 infection and restoration by vaccine-induced HIV-1-specific CD4⁺ T cells. *J. Exp. Med.* **200**, 701–712 (2004).
8. Kostense, S. *et al.* Persistent numbers of tetramer⁺ CD8⁺ T cells, but loss of interferon-γ⁺ HIV-specific T cells during progression to AIDS. *Blood* **99**, 2505–2511 (2002).
9. Kostense, S. *et al.* High viral burden in the presence of major HIV-specific CD8⁺ T cell expansions: evidence for impaired CTL effector function. *Eur. J. Immunol.* **31**, 677–686 (2001).
10. Goeptfert, P. A. *et al.* A significant number of human immunodeficiency virus epitope-specific cytotoxic T lymphocytes detected by tetramer binding do not produce gamma interferon. *J. Virol.* **74**, 10249–10255 (2000).
11. Appay, V. *et al.* HIV-specific CD8⁺ T cells produce antiviral cytokines but are impaired in cytolytic function. *J. Exp. Med.* **192**, 63–75 (2000).
12. Zhang, D. *et al.* Most antiviral CD8 T cells during chronic viral infection do not express high levels of perforin and are not directly cytotoxic. *Blood* **101**, 226–235 (2003).
13. Rosenberg, E. S. *et al.* Vigorous HIV-1-specific CD4⁺ T cell responses associated with control of viremia. *Science* **278**, 1447–1450 (1997).
14. Kiepiela, P. *et al.* Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* **432**, 769–775 (2004).
15. Steiner, K. *et al.* Enhanced expression of CTLA-4 (CD152) on CD4⁺ T cells in HIV infection. *Clin. Exp. Immunol.* **115**, 451–457 (1999).
16. Tsomides, T. J. *et al.* Naturally processed viral peptides recognized by cytotoxic T lymphocytes on cells chronically infected by human immunodeficiency virus type 1. *J. Exp. Med.* **180**, 1283–1293 (1994).
17. Yang, O. O. *et al.* Efficient lysis of human immunodeficiency virus type 1-infected cells by cytotoxic T lymphocytes. *J. Virol.* **70**, 5799–5806 (1996).
18. Loffredo, J. T. *et al.* Tat_{28–35}SL8-specific CD8⁺ T lymphocytes are more effective than Gag_{181–189}CM9-specific CD8⁺ T lymphocytes at suppressing simian immunodeficiency virus replication in a functional *in vitro* assay. *J. Virol.* **79**, 14986–14991 (2005).
19. Champagne, P. *et al.* Skewed maturation of memory HIV-specific CD8 T lymphocytes. *Nature* **410**, 106–111 (2001).
20. Appay, V. *et al.* Memory CD8⁺ T cells vary in differentiation phenotype in different persistent virus infections. *Nature Med.* **8**, 379–385 (2002).
21. Hess, C. *et al.* HIV-1 specific CD8⁺ T cells with an effector phenotype and control of viral replication. *Lancet* **363**, 863–866 (2004).
22. Day, C. L. & Walker, B. D. Progress in defining CD4 helper cell responses in chronic viral infections. *J. Exp. Med.* **198**, 1773–1777 (2003).
23. Trabattoni, D. *et al.* B7–H1 is up-regulated in HIV infection and is a novel surrogate marker of disease progression. *Blood* **101**, 2514–2520 (2003).
24. Altman, J. D. *et al.* Phenotypic analysis of antigen-specific T lymphocytes. *Science* **274**, 94–96 (1996).
25. Dorfman, D. M., Brown, J. A., Shahsafaei, A. & Freeman, G. J. Programmed death-1 (PD-1) is a marker of germinal center-associated T cells and angioimmunoblastic T cell lymphoma. *Am. J. Surg. Pathol.* **30**, 802–810 (2006).
26. Brown, J. A. *et al.* Blockade of programmed death-1 ligands on dendritic cells enhances T cell activation and cytokine production. *J. Immunol.* **170**, 1257–1266 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by a Royal Society postdoctoral fellowship (C.L.D.), the Harvard University Center for AIDS Research (D.E.K. and B.D.W.), grants from the Doris Duke Charitable Foundation (B.D.W.), the NIH (B.D.W., G.J.F., R.A., P.J.R.G.), the Howard Hughes Medical Institute (B.D.W.); a grant from the Foundation for the National Institutes of Health through the Grand Challenges in Global Health initiative (G.J.F. and R.A.), a contract from the NIH (B.D.W.), and the Mark and Lisa Schwartz Foundation. We thank S. Chetty, N. Ismail, N. Mkhwanazi, K. Nair, A. Piechocka-Trocha, A. Rathod and M. Vanderstock for technical assistance.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.D.W. (bwalker@partners.org).

Stoichiometry and turnover in single, functioning membrane protein complexes

Mark C. Leake^{1*}, Jennifer H. Chandler^{2*}, George H. Wadhams², Fan Bai¹, Richard M. Berry¹ & Judith P. Armitage²

Many essential cellular processes are carried out by complex biological machines located in the cell membrane. The bacterial flagellar motor is a large membrane-spanning protein complex that functions as an ion-driven rotary motor to propel cells through liquid media^{1–3}. Within the motor, MotB is a component of the stator that couples ion flow to torque generation and anchors the stator to the cell wall^{4,5}. Here we have investigated the protein stoichiometry, dynamics and turnover of MotB with single-molecule precision in functioning bacterial flagellar motors in *Escherichia coli*. We monitored motor function by rotation of a tethered cell body⁶, and simultaneously measured the number and dynamics of MotB molecules labelled with green fluorescent protein (GFP–MotB) in the motor by total internal reflection fluorescence microscopy. Counting fluorophores by the stepwise photobleaching of single GFP molecules showed that each motor contains ~22 copies of GFP–MotB, consistent with ~11 stators each containing two MotB molecules. We also observed a membrane pool of ~200 GFP–MotB molecules diffusing at ~0.008 $\mu\text{m}^2 \text{s}^{-1}$. Fluorescence recovery after photobleaching and fluorescence loss in photobleaching showed turnover of GFP–MotB between the membrane pool and motor with a rate constant of the order of 0.04 s^{-1} : the dwell time of a given stator in the motor is only ~0.5 min. This is the first direct measurement of the number and rapid turnover of protein subunits within a functioning molecular machine.

Over 30% of all proteins are integrated in biological membranes, where they carry out diverse essential cellular functions. Many membrane proteins function in multimeric complexes, and investigating their organization and dynamics is essential for understanding their function. GFP fusion constructs⁷ have been used with total internal reflection fluorescence (TIRF) microscopy to detect single interactions at the basal membrane *in vivo*⁸, and fluorescence recovery after photobleaching (FRAP) and fluorescence loss in photobleaching (FLIP) have been used to measure dynamics of membrane protein populations^{9–12}. Little is understood, however, about protein dynamics and turnover within individual complexes under natural conditions in living cells. Here we show that protein components of a functioning biological machine undergo rapid exchange with a freely diffusing pool in the cell membrane.

The bacterial flagellar motor is ideal for examining a single protein complex *in vivo*; rotation of the whole cell when the filament is attached to a surface (Fig. 1) is an instantaneous indicator of motor function. Previous research indicates that a motor has 8–16 stators^{4,5,13–15}, each containing two copies of MotB and four copies of MotA¹⁶. We replaced genomic *motB* in *E. coli* with *gfp-motB* to express GFP–MotB in wild-type amounts (see Methods). Polystyrene beads (0.75 μm in diameter) attached to filaments¹⁷ rotated three times slower than on the parental strain (Supplementary Fig. 3),

indicating some reduction in GFP–MotB function. We attached live cells to a coverslip either by the filament ('tethered') or by the cell body ('stuck') and observed them with TIRF or brightfield microscopy (Fig. 1a). TIRF images of tethered cells showed spots at the centre of cell rotation measured from brightfield images (Fig. 1c, Supplementary Videos 1 and 2). Similar spots were observed for stuck cells, with one or occasionally two spots visible. Spot size and

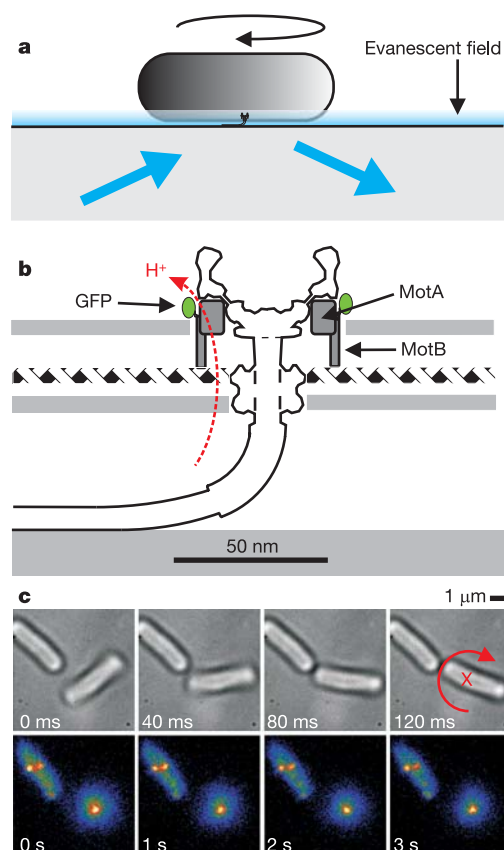


Figure 1 | TIRF microscopy of live GFP–MotB cells. **a**, **b**, Antibody-tethered cell rotation assay (**a**) and expansion of the motor structure (**b**). Coverslip and cell membranes are light grey, the cell wall is hatched, and TIRF illumination is blue-green. **c**, Consecutive brightfield (top) and TIRF (bottom) images showing a rotating tethered GFP–MotB cell and a nearby stuck cell. Rotation of the freely tethered cell is indicated in red. Two motor spots are visible in the stuck cell, whereas one is visible at the centre of rotation in the tethered cell.

¹Clarendon Laboratory, Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK. ²Microbiology Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK.

*These authors contributed equally to this work.

Table 1 | Photobleach parameters for different cellular components

Component*	Initial intensity (I_0)	Time constant (t_0)
Motor GFP-MotB	$103,000 \pm 3,300 \text{ counts s}^{-1}$	$44 \pm 1 \text{ s}$
Membrane GFP-MotB	$280 \pm 50 \text{ counts s}^{-1} \text{ pixel}^{-1}$	$269 \pm 8 \text{ s}$
Cell autofluorescence	$1,351 \pm 42 \text{ counts s}^{-1} \text{ pixel}^{-1}$	$62 \pm 1 \text{ s}$

*Data are from 134 ROIs in GFP-MotB cells containing motors (row 1), 27 ROIs in GFP-MotB cells containing no motors (row 2), and 32 ROIs in cells lacking GFP-MotB (row 3). Data are mean $\pm$ s.e.m.

number were consistent with a ring of GFP-MotB molecules bordering a rotor with a diameter of $\sim 50 \text{ nm}$ (refs 14, 18; Fig. 1b) and ~ 6 motors per cell¹⁹ (Supplementary Note 1).

We modelled fluorescence intensity (I) in 400-nm square regions of interest (ROIs) containing a single motor as a uniform background plus a gaussian spot of width 300 nm (motor width plus microscope point-spread function) and identified three background components: GFP-MotB in the membrane, nonspecific cellular autofluorescence and instrumental background (Supplementary Methods 2). Cytoplasmic GFP and autofluorescence were not expected to contribute because only the cell surface adjacent to the coverslip was illuminated. We estimated instrumental background from an empty ROI close to the cell in each image and autofluorescence from the parental strain. All cellular components showed roughly exponential photobleaching under TIRF illumination: $I(t) = I_0 \exp(-t/t_0)$ (Fig. 2a, Table 1 and Supplementary Fig. 4). The GFP membrane component bleached considerably more slowly than the motor, which we attribute to diffusive exchange with unbleached fluorophores from outside the TIRF field.

Total intensity (minus instrumental background) in ROIs containing motors photobleached in steps at roughly integer multiples of a unitary level, I_{GFP} , consistent with photobleaching of individual GFP

molecules^{20,21} (Fig. 2b, e). Separation into motor and background components introduced extra noise (Fig. 2a); therefore, we calculated I_{GFP} using total intensity rather than the motor component alone. Figure 2c and d shows the pairwise-distance distribution function (PDDF; Supplementary Methods 3) and its power spectrum, respectively, for the filtered curve 3 in Fig. 2b (refs 22, 23). The peak in the power spectrum indicates a unitary step size of $I_{\text{GFP}} \approx 5,400 \text{ counts}$ (ref. 23). To confirm that this corresponds to bleaching of one GFP molecule, we reduced the background fluorescence by prebleaching the cell, facilitating direct observation of successive photobleaching steps in the motor component of fluorescence (Fig. 2e and Supplementary Methods 4). The double-sized step in Fig. 2e presumably corresponds to two successive unitary photobleaches that were not resolved by the filtering algorithm; furthermore, no steps were detected in photobleaches of the parental strain. Step-wise photobleaching of single surface-immobilized GFP molecules²⁴ showed an average step size of $\sim 13,000 \text{ counts s}^{-1}$ in our microscope, after correcting for differences in laser power and exposure time (Fig. 2f). TIRF intensity falls exponentially with distance over $\sim 100 \text{ nm}$ (Supplementary Methods 5); thus, the average value of I_{GFP} is consistent with motors being $\sim 90 \text{ nm}$ from the coverslip.

Supplementary Fig. 5a and b show distributions of I_{GFP} and initial motor intensities (I_0^m), respectively, for 134 traces from different cells. The widths probably reflect different TIRF intensities, which varied about fourfold over the measured range of cell heights ($\sim 150 \text{ nm}$) owing to different distances from motor to coverslip. We estimated the total number of GFP-MotB molecules per motor by dividing I_0^m by I_{GFP} for each trace (Supplementary Fig. 5c). The reduced uncertainty of this estimate (22 ± 6 , $\pm$ s.d.) as compared with the ratio of peaks in Supplementary Fig. 5a and b (19 ± 8) reflects covariation of I_0^m and I_{GFP} , as expected for variations due to TIRF intensity. Dividing the initial membrane component of

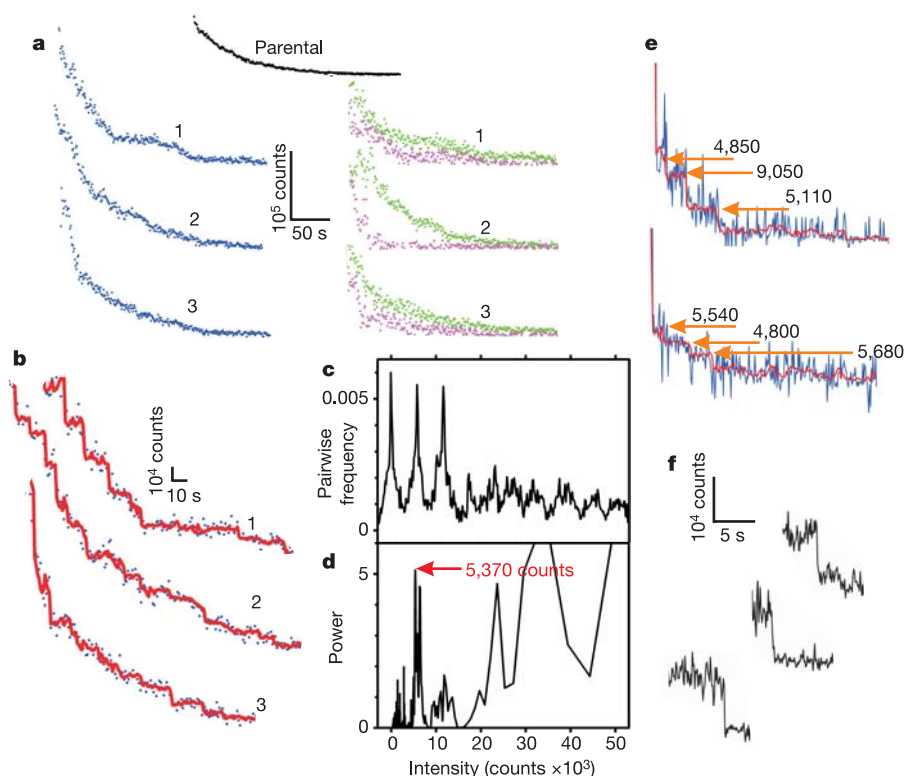


Figure 2 | TIRF photobleaching. **a**, Three photobleaches for regions centred on motors, showing total (blue), motor (magenta) and background (green) intensities, and average autofluorescence (black) from 32 parental cells lacking GFP. **b**, Expansions from traces in **a** with C-K-filtered traces overlaid (red). **c**, **d**, PDDF (**c**) and power spectrum of the PDDF (**d**) for filtered

intensity curve 3 in **a**; the unitary peak is indicated (arrow). **e**, Stepwise photobleaching of motors after prebleaching of the cell to reduce background. The motor component (blue), C-K-filtered trace (red), steps detected and step sizes are indicated (arrows). **f**, Stepwise photobleaching of surface-immobilized GFP molecules²⁴.

background intensity by I_{GFP} gave an average of 0.052 ± 0.022 molecules per pixel. We estimated the cell-surface area as $3,700 \pm 500$ pixels; thus, the total number of non-motor GFP–MotB molecules per cell is 190 ± 80 .

We investigated protein turnover between membrane and motor components using focused laser exposures of 0.5 s to photobleach all fluorophores in regions with a width of $\sim 1 \mu\text{m}$. Figure 3a shows TIRF images of a cell before and after bleaching of a region containing a motor. Fluorescence recovery (FRAP) of both motor and background components in the bleached region is visible, as is fluorescence loss ('one-shot' FLIP; Supplementary Methods 5) at the other end of the cell. FRAP and FLIP (Fig. 3b) of motor and membrane components respectively, averaged over 13–38 cells, are shown in Fig. 3c and d. Both components recovered over the course of a few minutes; motor recovery was slightly delayed as compared with membrane. Fluorescence loss occurred on the same timescale in both components.

Fitting all of the membrane FRAP data from ROIs lacking motors to a model of a diffusing GFP–MotB membrane pool gave a diffusion coefficient, D , of $0.0075 \pm 0.0013 \mu\text{m}^2 \text{s}^{-1}$ (Supplementary Methods 6). We obtained an independent, single-molecule estimate, $D = 0.0088 \pm 0.0026 \mu\text{m}^2 \text{s}^{-1}$, by tracking mobile fluorescent spots

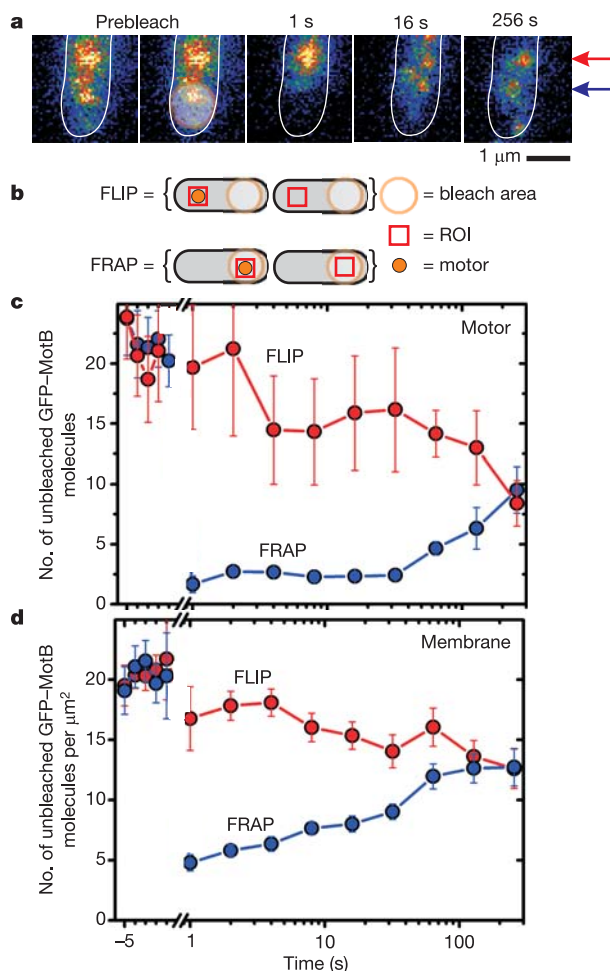


Figure 3 | Focused laser FRAP and 'one-shot' FLIP. **a**, Successive TIRF images of a GFP–MotB cell before and after bleaching. The prebleach images are identical; the laser focus is indicated (circle, right panel). Arrows indicate positions of two motors, showing FLIP (red) and FRAP (blue); the cell is outlined (white). **b**, Representation of FLIP and FRAP using ROIs with and without motors. **c**, **d**, Mean numbers of unbleached GFP–MotB molecules in motor (**c**) and membrane (**d**) components versus time. Data points were an average of 13 (motor, FLIP), 30 (motor, FRAP) and 38 (membrane) ROIs. Errors bars indicate 1 s.d.

that could occasionally be resolved after prebleaching (Supplementary Methods 6 and Supplementary Videos 3 and 4). The intensity of these spots was consistent with a GFP–MotB dimer (Supplementary Fig. 9). Fitting all of the FRAP and FLIP data to an extended diffusion model, incorporating binding and unbinding at the motor and bleaching in the TIRF field (Supplementary Methods 7), gave a dissociation rate at the motor of $0.04 \pm 0.02 \text{s}^{-1}$. The large uncertainty is a consequence of the small difference between recovery rates of motor and membrane components relative to noise. We also studied slow recovery after complete bleaching of whole cells (Supplementary Methods 8 and Supplementary Videos 5–7). After 90 min, fluorescence recovery ranged from 75% in buffer enriched with nutrients to 7% with protein synthesis blocked (attributable to GFP maturation⁷).

Fluorophore counting indicated that there are 22 ± 6 GFP–MotB molecules per motor. We confirmed the accuracy of the counting by testing it against simulated TIRF photobleach traces generated by the extended diffusion model (Supplementary Fig. 11). We also measured very low fluorescence anisotropy in motors after bleaching with polarized light (Supplementary Methods 9), indicating that GFP fluorophores in motors are free to rotate and therefore all are counted by the photobleaching method. Measurements of solubility and copurification with MotA suggest that MotB is stable in the cytoplasmic membrane only as a dimer in 1:2 stoichiometry with MotA¹⁶. This suggests that, on average, a motor has 11 stators, each with the composition MotA₄:MotB₂. Early work estimated 16 stators in a motor⁴, whereas subsequent studies estimated 8 stators⁵. We previously observed at least 11 discrete speed increments after induced expression of functional Mot proteins in a defective background¹⁵. Our photobleaching result also agrees with freeze-fracture electron microscopy showing 11–12 stators around the rotor¹⁴. The broad distributions in our counting estimates may include both natural variation in the number of stators per motor and noise due to background fluorescence, GFP blinking²⁰ and instrumental limitations. We measured a mobile membrane pool of ~ 200 GFP–MotB molecules with $D \approx 0.008 \mu\text{m}^2 \text{s}^{-1}$, which is $\sim 40\%$ smaller than the diffusion rate measured for free membrane proteins of comparable size to a GFP–MotB dimer²⁵, and only $\sim 20\%$ smaller than that for a MotA₄:(GFP–MotB)₂ stator, assuming that the Stokes radius r scales to the $1/3$ power with molecular weight and that D scales as $1/r$. A relatively high value of D agrees with other data suggesting that MotB in the membrane pool does not bind significantly to the cell wall²⁶. Eleven stators, each with a dissociation rate $\sim 0.04 \text{s}^{-1}$, gives an overall exchange of ~ 0.44 stators per s.

In summary, by replacing wild-type *motB* in the genome with *gfp-motB*, we expressed the fusion protein in natural quantities, enabling us to investigate protein stoichiometry and dynamics under physiological conditions. Our methods for counting the number of proteins in a membrane complex in living cells and observing their turnover using TIRF microscopy should be applicable to other membrane protein complexes. Unexpectedly, an anchored component, MotB²⁷, of a multiprotein complex was found to diffuse in the membrane and exchange rapidly with the motor, a finding that may alter the conventional 'static' view of molecular complexes; if proteins in the flagellar motor are undergoing dynamic exchange, this finding may well apply to other macromolecular complexes. It should be possible using similar methods to learn whether other membrane proteins also turnover rapidly and, if not, to investigate possible reasons for differences between complexes.

METHODS

***E. coli* strains and cell preparation.** A strain expressing GFP–MotB from the genome in normal MotB physiological quantities was constructed, and cells were prepared as described in Supplementary Methods 1.

Fluorescence microscopy. We used a home-built inverted TIRF microscope with an excitation wavelength of 488 nm (Supplementary Methods 5). Fluorescence emission was imaged at $\sim 50 \text{ nm}$ per pixel in frame-transfer mode at 1 Hz

(25 Hz for immobilized GFP molecules, 10 Hz for particle tracking) by a 128×128 -pixel, cooled, back-thinned electron-multiplying charge-coupled-device camera (iXon DV860-BI, Andor Technology).

Image acquisition and photobleaching. Images were sampled continuously for 300 s, resulting in >90% photobleaching within range of the TIRF field. For FRAP and FLIP experiments, single TIRF exposures were taken at intervals up to 256 s after bleaching with a focused laser spot for 0.5 s, centred either over a fluorescent spot of a putative motor (FRAP) or >1 μm from a motor (FLIP). Motor and membrane components were separated as described in the text for TIRF bleaches. Average curves were generated for FRAP and FLIP of both motor and membrane components; all intensity components were corrected for photobleaching (Supplementary Methods 5).

Data analysis and simulations. Continuous TIRF intensity data were filtered by using a Chung–Kennedy (C–K) algorithm²²; spatial frequency analysis of the pairwise intensity-difference histograms was used to determine the unitary step size²³; membrane diffusion was modelled by using Monte Carlo simulations for mobility of single GFP–MotB molecules; rate constants for the turnover process at the motor were evaluated by using a least-squares fit to FRAP and FLIP data (see Supplementary Information for details).

Received 21 June; accepted 24 July 2006.

Published online 13 September 2006.

- Macnab, R. M. in *Escherichia coli and Salmonella: Cellular and Molecular Biology* (ed. Neidhardt, F. C.) 123–145 (American Society for Microbiology, Washington DC, 1996).
- Berry, R. M. & Armitage, J. P. The bacterial flagella motor. *Adv. Microb. Physiol.* **41**, 291–337 (1999).
- Berg, H. C. The rotary motor of bacterial flagella. *Annu. Rev. Biochem.* **72**, 19–54 (2003).
- Block, S. M. & Berg, H. C. Successive incorporation of force-generating units in the bacterial rotary motor. *Nature* **309**, 470–472 (1984).
- Blair, D. F. & Berg, H. C. Restoration of torque in defective flagellar motors. *Science* **242**, 1678–1681 (1988).
- Silverman, M. & Simon, M. Flagellar rotation and the mechanism of bacterial motility. *Nature* **249**, 73–74 (1974).
- Tsien, R. Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509–544 (1998).
- Mashanov, G. I., Tacon, D., Peckham, M. & Molloy, J. E. The spatial and temporal dynamics of pleckstrin homology domain binding at the plasma membrane measured by imaging single molecules in live mouse myoblasts. *J. Biol. Chem.* **279**, 15274–15280 (2004).
- Mullineaux, C. W. & Sarcina, M. Probing the dynamics of photosynthetic membranes with fluorescence recovery after photobleaching. *Trends Plant Sci.* **7**, 237–240 (2002).
- Ray, N., Nenner, A., Mullineaux, C. W. & Robinson, C. Location and mobility of twin-arginine translocase subunits in the *Escherichia coli* plasma membrane. *J. Biol. Chem.* **280**, 17961–17968 (2005).
- Goodwin, J. S. & Kenworthy, A. K. Photobleaching approaches to investigate diffusional mobility and trafficking of Ras in living cells. *Methods* **37**, 154–164 (2005).
- Mullineaux, C. W., Nenner, A., Ray, N. & Robinson, C. Diffusion of green fluorescent protein in three cell environments in *Escherichia coli*. *J. Bacteriol.* **188**, 3442–3448 (2006).
- Berry, R. M., Turner, L. & Berg, H. C. Mechanical limits of bacterial flagellar motors probed by electrorotation. *Biophys. J.* **69**, 280–286 (1995).
- Khan, S., Dapice, M. & Reese, T. S. Effects of *mot* gene expression on the structure of the flagellar motor. *J. Mol. Biol.* **202**, 575–584 (1988).
- Reid, S. W. *et al.* The maximum number of torque-generating units in the flagellar motor of *Escherichia coli* is at least 11. *Proc. Natl Acad. Sci. USA* **103**, 8066–8071 (2006).
- Kojima, S. & Blair, D. F. Solubilization and purification of the MotA/MotB complex of *Escherichia coli*. *Biochemistry* **43**, 26–34 (2004).
- Sowa, Y. *et al.* Direct observation of steps in rotation of the bacterial flagellar motor. *Nature* **437**, 916–919 (2005).
- Sourjik, V. & Berg, H. C. Localization of components of the chemotaxis machinery of *Escherichia coli* using fluorescent protein fusions. *Mol. Microbiol.* **37**, 740–751 (2000).
- Turner, L., Ryu, W. S. & Berg, H. C. Real-time imaging of fluorescent flagellar filaments. *J. Bacteriol.* **182**, 2793–2801 (2000).
- Dickson, R. M., Cubitt, A. B., Tsien, R. Y. & Moerner, W. E. On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature* **388**, 355–358 (1997).
- Chu, S. Biology and polymer physics at the single-molecule level. *Phil. Trans. R. Soc. Lond. A* **361**, 689–698 (2003).
- Leake, M. C., Wilson, D., Gautel, M. & Simmons, R. M. The elasticity of single titin molecules using a two-bead optical tweezers assay. *Biophys. J.* **87**, 1112–1135 (2004).
- Svoboda, K., Schmidt, C. F., Schnapp, B. J. & Block, S. M. Direct observation of kinesin stepping by optical trapping interferometry. *Nature* **365**, 721–727 (1993).
- Mashanov, G. I., Tacon, D., Knight, A. E., Peckham, M. & Molloy, J. E. Visualizing single molecules inside living cells using total internal reflection fluorescence microscopy. *Methods* **29**, 142–152 (2003).
- Deich, J., Judd, E. M., McAdams, H. H. & Moerner, W. E. Visualization of the movement of single histidine kinase molecules in live *Caulobacter* cells. *Proc. Natl Acad. Sci. USA* **101**, 15921–15926 (2004).
- Van Way, S. M., Hosking, E. R., Braun, T. F. & Manson, M. D. Mot protein assembly into the bacterial flagellum: a model based on mutational analysis of the *motB* gene. *J. Mol. Biol.* **297**, 7–24 (2000).
- Demot, R. & Vanderleyden, J. The C-terminal sequence conservation between OMPA-related outer membrane proteins and MotB suggests a common function in both gram-positive and gram-negative bacteria, possibly in the interaction of these domains peptidoglycan. *Mol. Microbiol.* **12**, 333–336 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

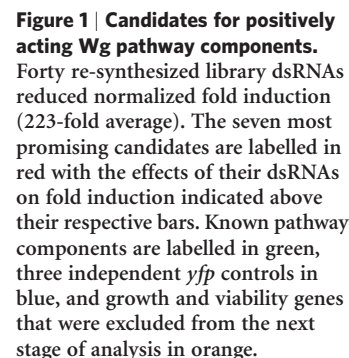
Acknowledgements We thank D. Blair for antibodies to flagellin and MotB. The research of M.C.L., J.H.C., R.M.B. and J.P.A. was supported by combined UK research councils via an Interdisciplinary Research Collaboration in Bionanotechnology (IRC), that of G.H.W. by the Biotechnology and Biological Sciences Research Council (BBSRC), and that of F.B. by a Clarendon Scholarship.

Author Contributions Fluorescence experiments were carried out by M.C.L. and J.H.C. in the laboratory of R.M.B.; strain construction was done by J.H.C. and G.H.W. in the laboratory of J.P.A.; data analysis was done by M.C.L., R.M.B. and G.H.W.; and simulations were carried out by F.B. and M.C.L. The experiment was designed by M.C.L., R.M.B. and J.P.A.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.P.A. (judith.armitage@bioch.ox.ac.uk).

RNA interference (RNAi) in both plants and animals is mediated by small RNAs of approximately 21–23 nucleotides in length for regulation of target gene expression at multiple levels through partial sequence complementarities^{1,2}. Combined with widespread genome sequencing, experimental use of RNAi has the potential to interrogate systematically all genes in a given organism with respect to a particular function^{3–9}. However, owing to a tolerance for mismatches and gaps in base-pairing with targets^{10–12}, small RNAs could have up to hundreds of potential target sequences in a genome^{13,14}, and some small RNAs in mammalian systems have been shown to affect the levels of many messenger RNAs besides their intended targets^{15,16}. The use of long double-stranded RNAs (dsRNAs) in *Drosophila*, where Dicer-mediated processing produces small RNAs inside cells, has been thought to reduce the probability of such ‘off-target effects’ (OTEs)⁵. Here we show, however, that OTEs mediated by short homology stretches within long dsRNAs are prevalent in *Drosophila*. We have performed a genome-wide RNAi screen for novel components of Wingless (Wg) signal transduction¹⁷ in *Drosophila* S2R+ cells, and found few, if any, legitimate candidates. Rather, many of the top candidates exert their effects on Wg response through OTEs on known pathway components or through promiscuous OTEs produced by tandem trinucleotide repeats present in many dsRNAs and genes. Genes containing such repeats are over-represented in candidate lists from published screens, suggesting that they represent a common class of false positives. Our results suggest simple measures to improve the reliability of genome-wide RNAi screens in *Drosophila* and other organisms.

We selected S2R+ cells in preference to several other cell lines (Supplementary Fig. 1) for a high-throughput RNAi screen for novel Wg pathway components, using a 96-well format and a transcriptionally responsive firefly luciferase reporter that contains seven consecutive copies of consensus TCF-binding sites (SuperTopFlash¹⁸; typically 50–400-fold induction). A control reporter constitutively expressing *Renilla* luciferase was used to normalize for transfection efficiency and to serve as a general indicator of the health of transfected cells⁸. A dsRNA library consisting of >20,000 dsRNAs targeting >95% of the annotated *Drosophila* transcriptional units¹⁹ was screened as pools of three independent dsRNAs. Two hundred and fifty-four pools in the initial screen were identified as reducing Wg-stimulated reporter activities below an arbitrary threshold (Supplementary Fig. 2). Individual dsRNAs within these pools were subsequently screened, with pooled dsRNAs as controls. Five known pathway components, *armadillo* (*arm*), *arrow* (*arr*), *legless* (*lgs*), *pangolin* (*pan*) and *pygopus* (*pygo*), were identified from these pools, validating the efficacy of our screening approach. Seventy dsRNAs that reproduced the effects of their parent pools were selected for further analysis, and these dsRNAs were synthesized *de novo* to ensure uniform yield and quality; their effects on fold induction (ratio of reporter activities in the presence and absence of Wg) spanned the entire spectrum of possible outcomes (Fig. 1 and Supplementary Fig. 3). Exclusion of dsRNAs previously reported to cause growth arrest or cell death¹⁹ left seven (*HDC04705*, *CG31374*, *CG32465*, *CG6834*, *l(1)G0003*, *CG8538* and *CG12993*) that reduced fold induction by more than 2.5 fold; cDNAs for six of these were readily available.



© 2006 Nature Publishing Group

To confirm the roles of these six genes in Wg signal transduction, additional segments of dsRNA covering the entire lengths of available cDNAs were tested. Notably, the only dsRNAs effective in inhibiting Wg response were those that overlapped with particular regions of the original dsRNAs from the library (Fig. 2a). The dsRNAs that were effective in S2R+ cells, however, were also effective in cl-8 cells (Fig. 2a), confirming that their effects are specific and are not restricted to a particular cell type. To test the possibility that ineffective dsRNAs were not efficiently targeting their cognate genes, expression constructs without untranslated regions (UTRs) were epitope-tagged for those genes with available full-length open reading frames (ORFs): *CG31374*, *CG6834*, *CG8538* and *CG12993*. As expected, neither UTR dsRNAs nor control *yfp* dsRNA affected the levels of proteins transiently overexpressed in S2R+ cells (Fig. 2b). All but one of the remaining 22 dsRNAs significantly reduced levels of their cognate proteins, including those that did not affect the Wg reporter assay. A series of mapping experiments further indicated that the Wg pathway effects of all seven library dsRNAs reside exclusively within 30–40-base-pair (bp) regions (Fig. 2c), suggesting that OTEs mediated by these shorter sequences might be responsible. Consistent with such a hypothesis, expression of the four candidates carrying translationally silent changes to render them resistant to dsRNA treatments failed to rescue Wg reporter activities under these dsRNA treatments (Supplementary Fig. 4).

The 30-bp effective sequence present in *CG32465* was analysed in finer detail by creation of six dsRNA pools, in each of which a group of five consecutive positions was replaced by a mixture of the three

base pairs other than that in the original sequence (Fig. 3a). Four dsRNA pools lost the inhibitory ability of the original dsRNA, thus identifying a 20-bp region containing the sequences required for this effect. In each of 20 additional dsRNA pools tested, a single base pair within the effective region was replaced, and a 16-bp region responsible for the OTE was identified by marked reductions in effectiveness of dsRNA pools. A BLAST homology search with this 16-bp sequence identified a perfect match to *arm* mRNA sequence. Notably, all the other six effective dsRNA sequences also contain distinct short homologies to *arm* (Supplementary Fig. 5a), indicating that their effects on Wg reporter activities may result from OTEs on *arm*.

To investigate this possibility, S2R+ cells treated with 30–40-bp dsRNAs (or a 248-bp dsRNA for *CG12993*) were subsequently transfected with a control or Wg expression construct. The levels of *arm* mRNA and Arm protein were reduced in stimulated (Wg-transfected; Fig. 3b) or unstimulated (control-transfected; Supplementary Fig. 6a) cells, consistent with action directed against *arm* expression. In addition, quantitative real-time RT-PCR showed that all seven dsRNAs reduced mRNA levels of the endogenous Wg signalling response target, *naked²⁰* (Supplementary Fig. 6b). Each of the six short homologous sequences from *arm* was also inserted immediately after the stop codon of the firefly luciferase ORF under the control of *actin5C* promoter to generate a series of reporter constructs (Supplementary Fig. 5b). All but *l(1)G0003* short dsRNA significantly reduced the activities of their respective reporters, indicating that the short *arm* homologies are sufficient to mediate the observed OTEs (Fig. 3c).

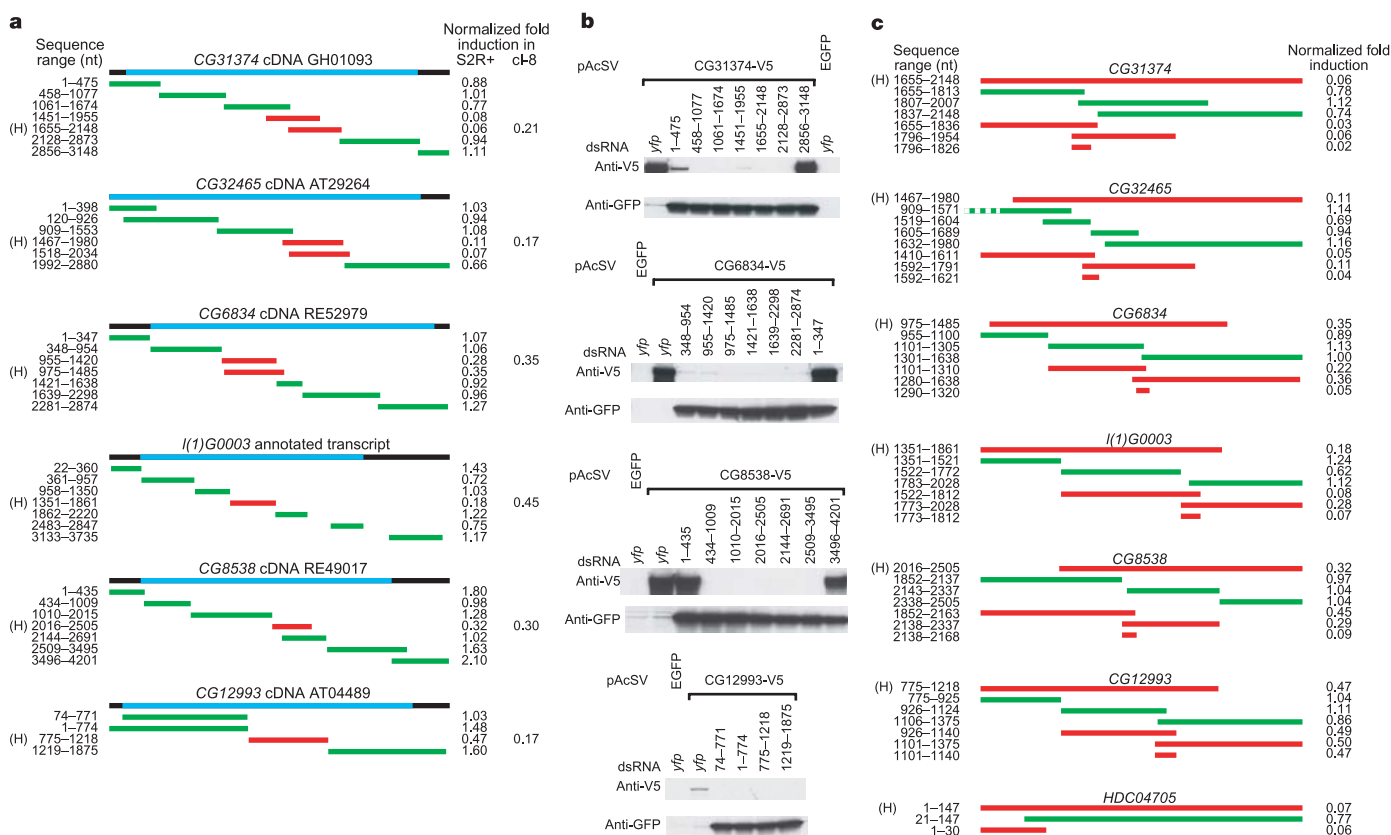


Figure 2 | Suppression of Wg pathway activity by the top seven candidate dsRNAs maps to short ~30–40-bp fragments. **a**, Only library dsRNAs and some overlapping dsRNAs suppressed Wg reporter activities. cDNAs (annotated transcript for *l(1)G0003*) are represented with ORFs in blue and UTRs in black. dsRNAs with effects similar to those of their corresponding library dsRNAs (marked by H) are represented by red lines; those having no effect or much weaker effects by green lines. The sequence range covered by each dsRNA relative to the respective template is shown on the left; the

normalized fold induction produced by each dsRNA is on the right. **b**, V5-tagged expression constructs and a control EGFP construct were transiently transfected into S2R+ cells together with the indicated dsRNAs. All dsRNAs except for *CG8538* (1–435) were effective in knocking down expression of their cognate proteins. **c**, The effective sequences within all seven library dsRNAs are further mapped to small regions of ~30–40 bp in length.

The clear OTEs of the top seven candidate dsRNAs prompted us to ask whether potential OTEs could also account for the effects of other candidate dsRNAs on Wg reporter activities. BLAST searches were performed for the 22 dsRNAs that reduced the normalized fold induction more than 1.5 fold (Fig. 1). The sequences of 19 dsRNAs contain short homologies (<30 bp) to known pathway components (Supplementary Table 1) for which RNAi produced significant inhibitory effects in the pilot experiment (*arm*, *arr*, *frizzled 2* (*fz2*), *lgs*, *pan* and *pygo*; see Supplementary Fig. 1). The three dsRNAs without BLAST-detectable homologies to these known components

were classified as ‘growth and viability’ genes in a previous study¹⁹, raising the question of whether their apparent effects on Wg reporter activities are the result of direct interference with Wg signalling response or indirect effects secondary to their cytotoxicity. Independent dsRNAs that do not overlap with library dsRNAs were tested for an additional seven candidates with readily available cDNAs. With the exception of *fizzy* (*fzy*), one of the growth and viability genes, these non-overlapping dsRNAs did not reproduce the inhibitory effects of their library dsRNAs in a Wg reporter assay (Supplementary Table 2). This behaviour is consistent with the action of the library dsRNAs on Wg reporter activities through OTEs on pathway components. We note that most primary candidates from a genome-wide screen for novel components of the Hedgehog (Hh) signalling pathway also behaved in a similar fashion, suggestive of OTEs (A.C. and P.A.B., data not shown).

The prevalent OTEs observed in our Wg screen prompted us to examine candidates from a similar screen published earlier²¹. Preliminary BLAST searches identified short homologies between many candidate dsRNAs and known pathway components (data not shown). The dsRNA of one such candidate, *CG5402*, seemed to target both *lgs* and *pygo* (Supplementary Fig. 7), consistent with the observed RNAi phenotype. The effective sequences of two other candidates, *CG4136* and *warts* (*wts*), were mapped to regions containing simple tandem repeats of the trinucleotide CAN (N indicates any base, Fig. 4a and Supplementary Fig. 8). A large proportion of reported negative pathway components contains such repeats; our observations suggest that the effects of these dsRNAs on uninduced reporter activities may depend on the duration of dsRNA treatment. Thus, whereas 2-day treatment mildly reduced basal reporter activities, 5-day treatment more dramatically produced the opposite effect, consistent with their proposed negative roles in Wg signal transduction. In contrast to SuperTopFlash activity, levels of the control *Renilla* luciferase reporter were reduced, and this effect was especially pronounced after 5 days of dsRNA treatment, leading to the apparent increase in normalized pathway activity. Thus, extending the time of dsRNA treatment, while presumably allowing for greater turnover of targeted protein, unfortunately also increases the chances of observing OTEs.

To examine further the effects of dsRNAs containing CAN repeats, artificial dsRNAs containing 13 repeats of CAA, CAG, CAC or CAU were produced and tested in the Wg reporter assay. Whereas the (CAC)₁₃ and (CAU)₁₃ dsRNAs produced only mild effects on basal reporter activities and on control *Renilla* luciferase activities, strong effects were produced by the (CAA)₁₃ and (CAG)₁₃ dsRNAs, thus essentially recapitulating the effects observed for *CG4136* and *wts* dsRNAs (Fig. 4b). Although the apparent Wg pathway effects of CAN-repeat-containing dsRNAs could, in part, occur through CAN-repeat-containing pathway components such as *dishevelled* (*dsh*), *shaggy* (*sgg*), or *neurexin* (*nej*), specific RNAi of these components does

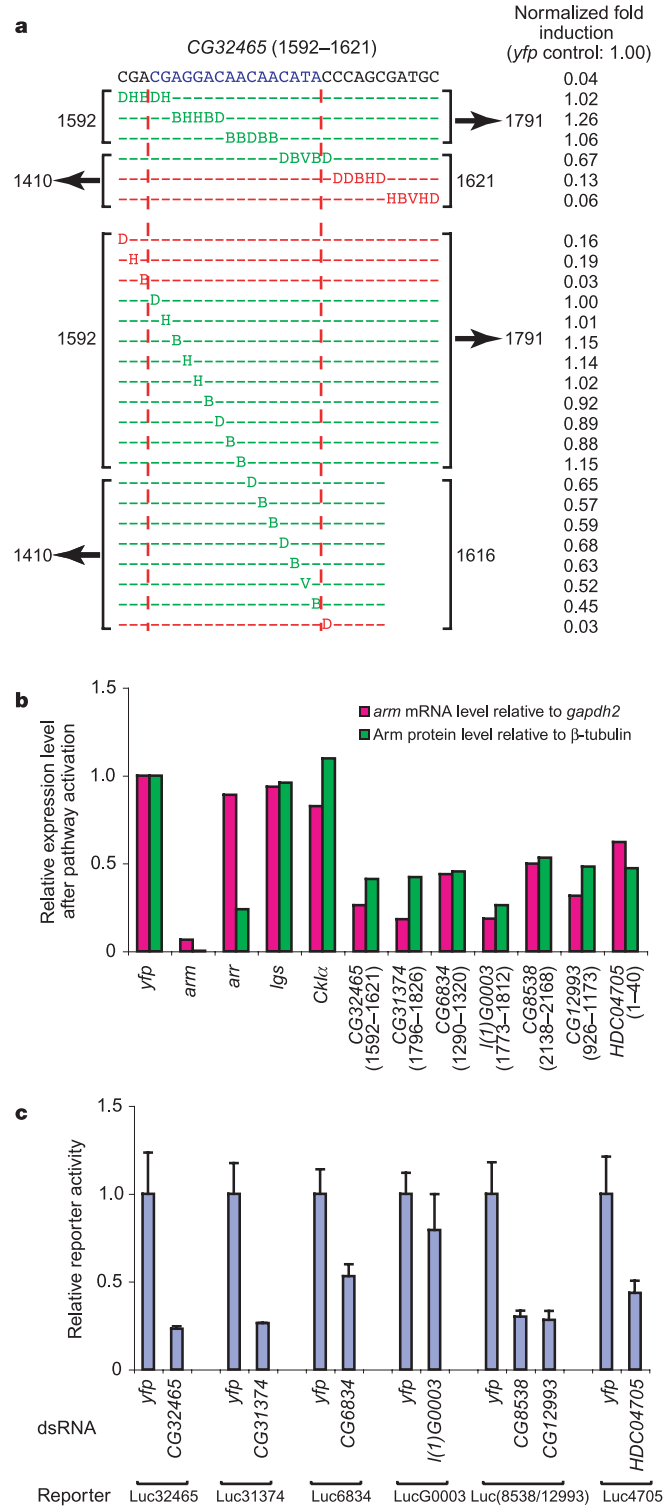


Figure 3 | Short homologies to *arm* account for the activity of top candidates. **a**, Sequence requirement for the Wg pathway effect of *CG32465* dsRNA. Mutational analysis of the 30-bp effective segment within *CG32465* initially identified a 20-bp region (with a series of 5-bp alterations) and subsequently 16 consecutive base pairs (with single-base-pair alterations) as essential for an inhibitory effect on Wg reporter activity. Altered bases within the DNA templates for each pool of dsRNAs are indicated according to the IUPAC code. Numbers flanking the sequences designate the end points of each dsRNA; sequences not shown are represented by arrows. Pools retaining or losing inhibitory effects are shown in red and green, respectively. **b**, Short dsRNAs (and a longer one for *CG12993*) reduced the levels of *arm* mRNA and Arm protein upon pathway activation. RNA levels were determined by quantitative real-time RT-PCR, and protein levels were quantified by NIH Image software analysis of a western blot. **c**, When each OTE luciferase reporter (see text) and a control *Renilla* luciferase reporter were co-transfected into S2R+ cells, five of the six short *arm* sequences were sufficient to mediate the RNAi effects of their respective short dsRNAs. Error bars represent standard deviations of triplicate experiments.

not reduce control *Renilla* luciferase activities (data not shown). The apparent cytotoxicity of CAN repeats thus appears to be due to disrupted expression of other CAN-repeat-containing genes or alternatively, to disruption of other unidentified cellular processes.

We also noted apparent effects of all four types of (CAN)₁₃ repeats on Hh induction of normalized reporter activities in Hh signalling assays in cl-8 cells (Fig. 4c), suggesting that the effects of CAN repeats may not be restricted to Wg reporter assays. Indeed, although CAN repeats are only present in ~5% of dsRNAs in the library, CAN-repeat-containing dsRNAs constitute up to 60% or more of the candidate genes listed in several published screens that used this library (Supplementary Table 3). Such enrichment for a large number and high proportion of dsRNAs containing CAN repeats suggests that candidate lists from these and other screens should be treated with caution.

Our genome-wide RNAi screen for positive components of the Wg signalling pathway in *Drosophila* S2R+ cells has identified few novel candidates. We have also reached a similar conclusion with this and other dsRNA libraries in screens for novel components of the Hh pathway (ref. 8 and A.C. and P.A.B., unpublished data). We do not

exclude the possibility that *bona fide* pathway components remain to be discovered^{22–24}, but our results suggest that the number of pathway components not yet identified in mutagenesis screens is limited. Moreover, our study unequivocally demonstrates that long dsRNAs, although remaining highly effective in targeting gene expression, also have the potential to generate phenotypically significant OTEs like those documented for short dsRNAs in mammalian systems^{15,16,25,26}. These OTEs are associated with unique sequences such as the *arm* homologues we noted in our Wg candidates or with repetitive simple sequences such as CAN repeats, which may produce pleiotropic effects through promiscuous targeting of related repeats in many gene products. A further complication of dsRNAs containing CAN repeats is that they can cause cytotoxicity, leading to identification of many genes as cell viability hits¹⁹ and consequent exclusion from further consideration as candidates in subsequent screens (see Supplementary Table 3).

Our experience suggests several simple measures that should produce more reliable lists of candidates in RNAi screens that use long dsRNAs. First, libraries should be designed to avoid sequences present in multiple genes, thus preventing the identification of false-positives through promiscuous OTEs. Second, phenotypic effects should be confirmed with more than one non-overlapping dsRNA for each candidate identified. Such testing with multiple dsRNAs would also have the benefit of reducing false positives that can arise due to noise inherent in screens of large numbers of items. Of course, RNAi-based screens at best merely provide a starting point in the identification and further mechanistic study of genetic elements with roles in a biological process of interest.

METHODS

Reagents. The antibodies used in this study are mouse anti-V5 monoclonal antibody (Invitrogen), rabbit anti-Myc antibody A-14 (Santa Cruz Biotechnology), mouse anti-Arm monoclonal antibody N2 7A1, anti-β-Tubulin monoclonal antibody E7 (Developmental Studies Hybridoma Bank), and rabbit anti-GFP antibody (Molecular Probes). Rabbit anti-CG31374 antibodies were generated against AA(1–147) (Spring Valley Lab). The luciferase activities in the cell lysates were measured with Dual-Luciferase Reporter Assay System or Dual-Glo Luciferase Assay System (Promega). Quantitative real-time PCR was performed with iQ SYBR green reagents (Bio-Rad). *Drosophila* cells were transfected with Effectene (Qiagen).

Wingless reporter assay. S2R+ cells were seeded in 96-well (initial screen) or 24-well (secondary screen and all subsequent experiments) plates on day 0. On day 1, six wells of cells were transfected with a single transfection mix containing SuperTopFlash reporter, control *Renilla* luciferase reporter, and various dsRNAs as indicated. On day 2, Wg-expressing and control S2 cells were each added to three wells of transfected S2R+ cells. The luciferase activities of the cell lysates were measured with a BMG FLUOstar OPTIMA luminometer on day 3.

Additional methods can be found in Supplementary Information.

Received 10 May; accepted 17 August 2006.

Published online 10 September 2006.

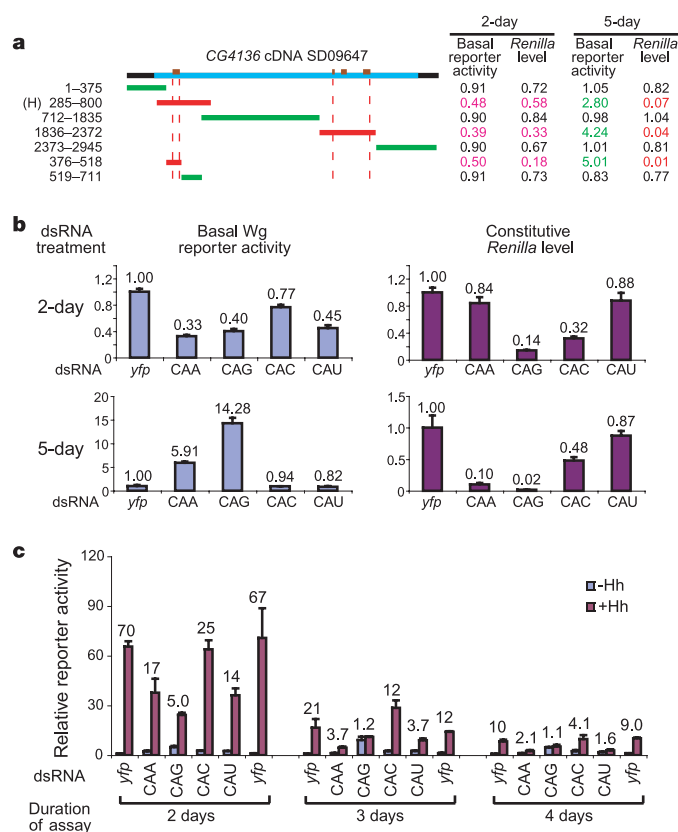


Figure 4 | Effects of CAN-repeat-containing dsRNAs in Wg and Hh reporter assays. **a**, The effective sequence of CG4136 dsRNA contains multiple stretches of CAN repeats (defined as ≥6 CAN repeats and shown as short brown lines above the cDNA schematic). Two-day and five-day dsRNA treatments have opposite effects on normalized basal reporter activity, but five-day treatment greatly enhanced the reduction of control *Renilla* luciferase reporter activity. **b**, (CAA)₁₃ and (CAG)₁₃ dsRNAs exhibited marked effects on normalized basal reporter activities and control *Renilla* luciferase activities, reminiscent of the effects observed with CG4136 dsRNA. Error bars represent standard deviations of triplicate experiments. **c**, All four (CAN)₁₃ dsRNAs increased basal Hh reporter activities in cl-8 cells, whereas (CAA)₁₃, (CAG)₁₃ and (CAU)₁₃ dsRNAs reduced the reporter activities upon Hh stimulation, resulting in an apparent decrease of fold induction (shown above each pair of bars). Note that the longer the treatment was, the less responsive the cells appeared. Error bars represent standard deviations of triplicate experiments.

- Valencia-Sanchez, M. A., Liu, J., Hannon, G. J. & Parker, R. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev.* **20**, 515–524 (2006).
- Zamore, P. D. & Haley, B. Ribo-gnome: the big world of small RNAs. *Science* **309**, 1519–1524 (2005).
- Moffat, J. & Sabatini, D. M. Building mammalian signalling pathways with RNAi screens. *Nature Rev. Mol. Cell Biol.* **7**, 177–187 (2006).
- Armknicht, S. et al. High-throughput RNA interference screens in *Drosophila* tissue culture cells. *Methods Enzymol.* **392**, 55–73 (2005).
- Echeverri, C. J. & Perrimon, N. High-throughput RNAi screening in cultured cells: a user's guide. *Nature Rev. Genet.* **7**, 373–384 (2006).
- Fraser, A. G. et al. Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408**, 325–330 (2000).
- Berns, K. et al. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431–437 (2004).
- Lum, L. et al. Identification of Hedgehog pathway components by RNAi in *Drosophila* cultured cells. *Science* **299**, 2039–2045 (2003).
- Paddison, P. J. et al. A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**, 427–431 (2004).
- Doench, J. G. & Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).

11. Haley, B. & Zamore, P. D. Kinetic analysis of the RNAi enzyme complex. *Nature Struct. Mol. Biol.* **11**, 599–606 (2004).
12. Brennecke, J., Stark, A., Russell, R. B. & Cohen, S. M. Principles of microRNA-target recognition. *PLoS Biol.* **3**, e85 (2005).
13. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
14. Lim, L. P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769–773 (2005).
15. Birmingham, A. *et al.* 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nature Methods* **3**, 199–204 (2006).
16. Jackson, A. L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nature Biotechnol.* **21**, 635–637 (2003).
17. Logan, C. Y. & Nusse, R. The Wnt signaling pathway in development and disease. *Annu. Rev. Cell Dev. Biol.* **20**, 781–810 (2004).
18. Veeman, M. T., Slusarski, D. C., Kaykas, A., Louie, S. H. & Moon, R. T. Zebrafish prickles, a modulator of noncanonical Wnt/Fz signaling, regulates gastrulation movements. *Curr. Biol.* **13**, 680–685 (2003).
19. Boutros, M. *et al.* Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* **303**, 832–835 (2004).
20. Zeng, W. *et al.* *naked cuticle* encodes an inducible antagonist of Wnt signalling. *Nature* **403**, 789–795 (2000).
21. DasGupta, R., Kaykas, A., Moon, R. T. & Perrimon, N. Functional genomic analysis of the Wnt-wingless signaling pathway. *Science* **308**, 826–833 (2005).
22. Mosimann, C., Hausmann, G. & Basler, K. Parafibromin/Hyrax activates Wnt/Wg target gene transcription by direct association with β -catenin/Armadillo. *Cell* **125**, 327–341 (2006).
23. Banziger, C. *et al.* Wntless, a conserved membrane protein dedicated to the secretion of Wnt proteins from signaling cells. *Cell* **125**, 509–522 (2006).
24. Bartscherer, K., Pelte, N., Ingelfinger, D. & Boutros, M. Secretion of Wnt ligands requires Evi, a conserved transmembrane protein. *Cell* **125**, 523–533 (2006).
25. Jackson, A. L. & Linsley, P. S. Noise amidst the silence: off-target effects of siRNAs? *Trends Genet.* **20**, 521–524 (2004).
26. Lin, X. *et al.* siRNA-mediated off-target gene silencing triggered by a 7 nt complementation. *Nucleic Acids Res.* **33**, 4527–4535 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank N. Perrimon and colleagues for their role in producing the dsRNA library, S. Yao for initial assistance in the Hh screen, and R. Gong for technical assistance. We also thank all members of the Beachy laboratory, R. Green and S. Dorner for helpful discussions, and G. Seydoux and J. Taipale for critical comments on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.A.B. (pbeachy@stanford.edu).

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

Nobody bats an eye when a student in the United States borrows \$100,000 to pay for a law degree, MBA or MD. Although that burden appears intimidating, the salaries these professions eventually pay make six-figure debts seem a good investment in the future.

Not so for scientists. Stipends for postdoctoral fellows in the United States, hovering around \$35,000–45,000, are often less than those earned by some professionals with only an undergraduate education. When loan repayments kick in, young scientists feel the squeeze — especially as they cannot expect a better salary until they achieve tenure, sometimes a decade down the road.

This financial burden could be one of the reasons why the number of US-born students pursuing research careers has declined over the past decade. And it could explain why scientific research hasn't succeeded in attracting many minorities. Why go through the education and face a career saddled with long odds and large debt, when a professional degree could yield double or triple the salary — and without the prospect of never-ending postdocs?

Comparing the United States with England is interesting. Graduates leaving university with large debts due to student loans are a more recent phenomenon in a country that used to have free tuition in higher education and grants for living expenses. A recent report from Research Councils UK shows that, since loans were introduced, fewer students have been interested in pursuing relatively poorly paid research careers.

Attempts to mitigate the problem are being made in the United States. Last week the National Institutes of Health called for applications for its annual programmes that offer help with student-loan repayments of up to \$35,000 a year to individuals working on government-funded research. These payments are a welcome start, but are very limited in the research fields they cover. Programmes with a wider remit might help turn around the decline. And if companies began to offer assistance with repayments, that might make science as attractive as business management and law.

Paul Smaglik, *Naturejobs* editor

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
Natureevents:
Claudia Paulsen Young (+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com
France/Switzerland/Belgium:
Muriel Lestringuez (4994)

UK/Ireland/Italy/RoW:
Loredana Milanese (4944)
Nils Moeller (4953)
Scandinavia/Spain/Portugal:
Evelina Rubio-Morgan (4973)
Germany/Austria/The Netherlands:
Reya Silao (4970)
Online Job Postings:
Matthew Ward (+44 (0) 20 7014 4059)

European Satellite Office
Germany: Patrick Phelan
Tel: +49 89 54 90 57 11
Fax: +49 89 54 90 57 20
e-mail: p.phelan@nature.com

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs web development: Tom Hancock
Naturejobs online production:
Catherine Alexander

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building, 2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com

UP AND COMING



L. SNIDER/PHOTO IMAGES/CORBIS

Of the thousands of visitors who come to New York each year, very few will look beyond the humming megatropolis of the city. But north, away from the urban sprawl, lies a network of distinctive cities and communities loosely termed 'upstate New York'. The region, more properly described as New York state, is known for its snowy, sometimes lengthy, winters, spectacular autumns and temperate summers.

Compared with the rest of the United States, the population of upstate New York is growing at a very sedate pace. Between 1970 and 2000, the national population growth was 38%, whereas upstate New York managed just 2%, says Richard Deitz, senior economist with the Federal Reserve Bank of New York in Buffalo. This reflects a general decline in manufacturing in the area and a general desire to move to the balmy south and east parts of the state. But at the same time, jobs in the biomedical sciences are diversifying and

Beyond the urban chaos of New York City lies a tranquil state with abundant career opportunities, says Ricki Lewis.

The East Campus of the University at Albany is home to a cancer genomic centre as well as biotech companies.

multiplying, making the area an attractive place to live for researchers tired of the urban bustle.

"A common view of upstate New York is that it has a lot of snow and nothing going on," says Bruce Holm, senior vice-provost of the University at Buffalo. But recently forged links between the region's research centres, industries, government labs and clinical facilities have created hundreds of jobs. In January 2006, for example, state governor George Pataki and state senate majority leader Joseph Bruno announced a \$200-million biotechnology and biomedicine research initiative, which is expected to generate triple that amount from external sources. The upstate biosciences sector certainly has room for growth: the New York Biotechnology Association has 130 members in New York City, 22 in the Hudson Valley and just 30 in the rest of the state.

Home sweet home

But one of upstate New York's great attractions is that it is simply a nice place to live. Affordable housing, excellent schools and natural beauty make it easier to recruit, says Gene Schuler, chair of Bioconnex, an organization that guides the development of biotechnology in Albany and its environs. "It's a wonderful place to raise kids," agrees Christoph Hergersberg, global technology leader of biosciences at the General Electric Global Research Center in Niskayuna, "although not locking doors took some getting used to after living in Manhattan, San Francisco and Munich".

One area that is attracting attention is the capital district, which encompasses Schenectady, Troy and the state capital Albany. This region has reinvented itself as 'technology valley', and is capitalizing on resident talent and drawing in newcomers. A decade ago, academic institutions in the area worked largely independently, Schuler says. Bioconnex was created in



part to get individuals from different institutions to talk to each other and share resources. For example, one collaboration of scientists from three centres recently won a programme project grant in biodefence from the National Institute of Allergy and Infectious Diseases to develop and test antimicrobials for drug-resistant strains of anthrax and plague.

Another example of the increasing links is the fledgling Alden March Bioethics Institute, headquartered at Albany Medical College. Its faculty members hail from 14 institutions, including the Rensselaer Polytechnic Institute (RPI), the Wadsworth Center and the Albany Law School. Glenn McGee, the institute's director, doesn't regret relocating from his position at the University of Pennsylvania. "The state capital is where bioethics happens," he says. "Physician-assisted suicide, stem cells, newborn screening and rules about reproduction have all shifted from the federal arena to the state."

General Electric has been a hub of scientific research in Schenectady since the General Electric Company was founded in the city in 1892. But its acquisition of Amersham Biosciences, now part of GE Healthcare, in 2003 created positions for biologists, chemists and computational scientists. When Hergersberg came to the company, the major draw, he says, was the synergy among the area's scientific employers. When both members of a couple are PhDs, scientists often turn down jobs when a significant other can't find work, Hergersberg notes. "That doesn't happen here."

Land of opportunity

The technology valley has also benefited from Gen*NY*sis (Generating Employment Through New York State Science), a biotechnology economic development programme established in 2003. The Center for Bioengineering and Medicine at the RPI is a new \$22.5-million arm of Gen*NY*sis, which should result in more high-paying, high-skilled jobs in the region, according to RPI president Shirley Ann Jackson. Another Gen*NY*sis facility is the Center for Excellence in Cancer Genomics on the East Campus of the University at Albany. This campus also houses companies such as Albany Molecular Research and Regeneron Pharmaceuticals, and non-profit organizations such as the state's Department of Environmental Conservation.



Both Laura Haynes (top) and Candace Johnson relish the academic interactions and energy at their institutes in upstate New York.

The Trudeau Institute focuses on infectious diseases and cancer research.



Benefits of living in technology valley abound. Sperling's Best Places, which rates US cities in various categories, recently ranked the region the "least stressful large metropolitan area" out of 100 regions featured. New York City ranked a stressful number five.

If the capital region is proving successful for technology, then the Buffalo region is emerging as a powerful centre for translational research. The area has a strong legacy of medical innovation: implantable pacemakers, Nicorette gum and photodynamic therapy are all examples of advances that hail from Buffalo, says Holm. He adds that recently the region has been on a massive recruitment drive to get more faculty members. This is largely the result of two new facilities: the Center of Excellence in Bioinformatics and Life Sciences at the University at Buffalo, and the Center for Genetics and Pharmacology at the Roswell Park Cancer Institute.

The Roswell Park institute of one of the 20 centres that make up the National Comprehensive Cancer Network in the United States, and it is the only comprehensive cancer centre in upstate New York. It underwent a renaissance in 1999, with an influx of new people, says Candace Johnson, the centre's senior vice-president for translational research. Since then, the number of positions for researchers and physicians and their staff has jumped from 2,100 to 2,700. The institute is recruiting heavily in translational research, Johnson says: "The energy here is palpable — and it's a fun place to be."

Boom time

Hiring is also booming an hour's drive from Buffalo, at the University of Rochester. Since 1996, the university's medical centre has recruited 500 new faculty members and staff, and doubled its funding from the National Institutes of Health. Multidisciplinary research centres are dedicated to a range of work, including musculoskeletal research, ageing and developmental biology. The Center for Future Health, for example, is taking a multidisciplinary approach to designing smart-sensor and information technologies that will allow people to manage health problems in their homes as much as they can.

For life scientists interested in basic research in infectious disease and with a passion for the outdoors, the Trudeau Institute is enticing. Located by Lake Saranac, the facility was once a tuberculosis sanatorium, where people from stifling New York City went for a 'rest cure'.

Today the private, not-for-profit institute employs 45 scientists and 110 support staff. There are some 20–30 postdocs, says faculty member Laura Haynes. The centre's research encompasses cancer, immunology and infectious diseases such as tuberculosis. Haynes compares life at Trudeau to an academic department. "It is big enough for a critical mass, and we interact quite a bit," she says.

For scientists who cherish nature, crave a slightly slower pace, and have no fear of snow, upstate New York may be a good option. Prospective researchers just need to get past the weather stigma, Holm says. "Once we break down the stereotype, point out the great work that has been and can be done here, and the affordable housing and great schools," he says, "the barriers just melt away."

Ricki Lewis is a freelance science writer in Albany, New York.

MOVERS

**Arthur Ellis, vice-chancellor for research,
University of California, San Diego**



2002-06: Director, chemistry division, National Science Foundation

1986-2006: Meloche-Bascom professor of chemistry, University of Wisconsin-Madison

1977-86: Assistant, associate, and professor of chemistry, University of Wisconsin-Madison

Chemistry played a major role in Arthur Ellis's life from an early age. As a child he could regularly be found experimenting with his chemistry set, but it was his high-school and college teachers that he credits as his true source of inspiration. As a result, Ellis's career has combined a zeal for chemistry with a desire to see the subject promoted and well taught.

Following his degree in chemistry, Ellis began his graduate career at the Massachusetts Institute of Technology in Cambridge, where he set about trying to build better solar panels by optimizing semiconductors and electrolytes. His graduate work earned Ellis the first of nine patents — and the luxury of going straight from graduate training into an assistant professorship at the University of Wisconsin in Madison. There his passion for teaching his subject quickly came to the fore, as he sought ways to bring the latest research to life in the classroom. In one demonstration, for instance, he levitated magnets using liquid nitrogen-cooled superconductors. "We can't have a static curriculum when research is so dynamic," he says.

Ellis's strong feelings about research and education, particularly the state of US chemistry training, finally led him to get involved in science policy. "After many years happily conducting my research and teaching, all the vectors lined up to do something different," he says. As a result, Ellis became director of the chemistry division at the National Science Foundation (NSF) in 2002.

Originally due to stay at the NSF for two years, Ellis ended up holding the position for four. While there, he helped to set up a number of initiatives. These included Chemical Bonding Centers, which support long-term projects intended to transform chemistry research; Discovery Corps Fellowships, which support nontraditional postdocs and sabbaticals directed at service-oriented projects; and undergraduate research collaboratives, which enable college students to gain experience in chemical research.

This month, Ellis arrived at the University of California, San Diego, to be its vice-chancellor for research. The university was keen to make use of Ellis's wide-ranging experience, says Marsha Chandler, its senior vice-chancellor for academic affairs. Ellis thinks his NSF experience will help him align his budget with university priorities. "I want to help create conditions that allow innovative ideas to bubble up from the campus and to take the university to the next level of research and education," he says. ■

Virginia Gewin

BRICKS & MORTAR

The mouse house

Early next year, Canada will open the doors on what it hopes will become a one-stop shop for mouse geneticists. The Toronto Centre for Phenogenomics (TCP) will house a range of imaging instruments, Canada's largest mouse colony and a cryobank, which together should make it one of the top locations for studying mouse models of human disease.

Affectionately nicknamed the mouse house, the Can\$63-million (US\$56-million) building will play host to some 300 researchers. They will generate mutant mice and work on phenotyping, behavioural analysis, imaging pathology and the cryopreservation of tissues. The human diseases that can be investigated using the mouse models range from diabetes and cancer to childhood conditions and women's health issues.

The TCP will become the new home to the mouse strains currently housed at its four sponsors: Mount Sinai Hospital, the Hospital for Sick Children, St Michael's Hospital and the University Health Network. This will free up space at the hospitals for other research.

But few of the researchers at these institutes will follow their rodent charges to the Toronto facility. Instead, the TCP aims to be a hub

of tools, technologies and services — previously available only on a piecemeal basis — and will house up to 200,000 mice.

At the TCP's core lies the Mouse Imaging Centre. Sealed in a room lined with 130 tonnes of steel is a 7.3-tesla magnetic resonance imaging machine that has been custom developed to image up to 19 live mice at a time. It also features optical projection tomography machines, which will allow researchers to create high-resolution three-dimensional images of transparent tissues, such as mouse embryos.

And at the Canadian Mouse Mutant Repository, sperm, ovaries, embryos and other non-germline tissue DNA will be held in cryopreservation. Colin McKelvie, who will oversee the repository and is currently the TCP's interim chief executive, is enthusiastic about the ways in which the tissue bank will help researchers. As well as being a useful resource, it should allow researchers to participate in international initiatives, he says.

With its state-of-the-art facilities, the opening of the TCP should have its investigators grinning like cats, as the first mice move in. ■

Hannah Hoag

GRADUATE JOURNAL

Farewell to the hive

I'm fortunate to have a supervisor who encourages her students to attend conferences. For me, conferences epitomize the idea of a scientific community where ideas are shared and debated. By presenting my work and gradually getting to know people, I feel increasingly at home in the community — or, to borrow a term from subjects of my research, the hive. When the hive gathers, anything can happen, from heated disputes between former pals to collaborations between strange bedfellows.

But for some of us, the hive is not for ever. At a conference last month, I faced the inevitable questions about my future plans. With graduation in sight, this was the ideal time to fish for postdoc positions. But instead I found myself telling colleagues about my intention to pursue a career outside research. I was met with everything from encouragement to awkward silences and downright incredulity, as if a life outside the hive could not be imagined, much less endured.

Leaving the hive is scary. Not many scientists will admit to considering alternatives, perhaps for fear of seeming less passionate about science. The conference atmosphere reminded me about my fascination for science. But it also encouraged me to look for a job where the things that I love about conferences — broad perspectives, interaction and communication — are everyday fare. ■

Katja Bargum is a graduate student in the Department of Biological and Environmental Sciences at the University of Helsinki, Finland.



Faculty Positions in Molecular Recognition and Bioinformatics

THE UNIVERSITY AT BUFFALO'S STRATEGIC PLANNING PROCESS, *UB 2020*, HAS IDENTIFIED BIOINFORMATICS AND MOLECULAR RECOGNITION IN BIOLOGICAL SYSTEMS AS AREAS WHERE MORE THAN 30 NEW HIRES ARE PROJECTED, WITH A SIGNIFICANT INVESTMENT OF RESOURCES, OVER THE NEXT THREE YEARS.



POSITION DESCRIPTIONS

In the first phase of our multi-year recruiting plan, several positions are available immediately at all ranks in the Schools of Medicine and Biomedical Sciences, Dental Medicine, Arts and Sciences, Pharmacy, and Engineering at the University at Buffalo, The State University of New York, with concomitant appointment in its New York State Center of Excellence in Bioinformatics and Life Sciences.

Successful candidates will have research interests in Bioinformatics and/or Molecular Recognition, in areas such as computational biology, genomics, molecular recognition in biology, signal transduction, and analysis of biological networks using experimental approaches ranging from *in silico* through *in vivo*.

Successful applicants will receive tenured or tenure-track appointments in departments appropriate to their interests. These include, but are not restricted to, Biochemistry, Biological Sciences, Chemical and Biological Engineering, Microbiology and Immunology, Oral Biology, Pharmacology and Therapeutics, Physiology and Biophysics, and Structural Biology. A history of collaborative research will be a key consideration in evaluation of applicants for senior positions, as will their potential to provide academic leadership and thereby facilitate future recruitments. Additional information on departmental programs may be obtained on specific home pages, accessible via www.buffalo.edu, or at www.bioinformatics.buffalo.edu.

STATE-OF-THE-ART FACILITIES

Ample laboratory space exists in the new Center of Excellence building, which opened in June 2006 and is situated adjacent to the Hauptman-Woodward Medical Research Institute and the Roswell Park Cancer Institute's Center for Genetics and Pharmacology in the Buffalo Niagara Medical Campus. When fully staffed, these three facilities will be home to over 500 biomedical scientists. The Center itself maintains a modern, state-of-the-art experimental infrastructure in biomedical sciences, and houses a major computational facility and high-quality confocal microscopy. Cutting-edge genomics and proteomics facilities are located immediately adjacent to the Center, as well as on the two local campuses of the University at Buffalo.

CANDIDATE REQUIREMENTS

Applicants should have Ph.D., M.D., D.V.M., D.D.S., or equivalent degree with postdoctoral and professional experience appropriate to the rank of appointment, e.g., a minimum of two years of postdoctoral experience for the rank of Assistant Professor. Research experience should be in molecular, cellular, genetic and/or computational approaches to signaling networks, molecular/microbial pathogenesis, development and differentiation or the basis of human disease. A well-established research program is essential for established candidates. Duties include teaching undergraduate, graduate and professional students, research training of students and postdoctoral fellows, and conduct of biomedical research.

APPLICATIONS

Applications in electronic format should include a C.V., brief statement of research interests, and the names and email addresses of three references, and should be directed to Bioinformatics/Molecular Recognition Search at ub-mrbs@buffalo.edu. Consideration of applications will continue until all available positions are filled.



University at Buffalo *The State University of New York*



Trudeau Institute, Inc. is interested in recruiting postdoctoral fellows to carry out research in murine models of immunology and infectious disease.

The research programs at Trudeau Institute include:

- Pulmonary immunity to viral and bacterial pathogens (including influenza, tuberculosis and plague)
- Immunity to protozoan and helminth infections
- Understanding immunologic memory
- Inflammation, Autoimmunity and Sepsis
- Cancer immunology
- Mucosal immunology
- Effects of aging on the immune system
- Novel approaches for vaccine development

In a 2005 survey by The Scientist Magazine, Trudeau Institute was ranked as one of the top five places to work among U.S. academic and non-commercial organizations. Also in 2005, The Institute was ranked fifth in the Best Places to Work for Postdocs among US Institutions.

Trudeau Institute is located in the heart of northern New York State's Adirondack Mountains offers competitive salaries, affordable housing, a full complement of benefits and on-site childcare.

Interested candidates should send a CV, cover letter, a brief statement of research interests and three references (preferably with an e-mail address included) to

Amy Richardson, Human Resource Manager, Trudeau Institute, Inc.,
154 Algonquin Avenue, Saranac Lake, NY 12983
or arichardson@trudeauinstitute.org

Further details may be found at www.trudeauinstitute.org



SP88485R



Scientists and Postdoctoral Fellows



The Ordway Research Institute, Inc., a biomedical research institute located in Albany, New York is recruiting funded scientists and postdoctoral fellows to join scientific research teams in Cancer, Cardiovascular Sciences (focus on fatty acid transport) and Infectious Diseases (focus on anthrax and tuberculosis).

Ordway is housed in a three-year old state-of-the-art research facility designed to facilitate collaboration and translational research. Ordway scientists enjoy the benefit of being located in New York's Tech Valley, with prestigious academic institutions and a growing biotech industry.

Salaries commensurate with experience. Excellent benefits. Competitive support for Postdoctoral Fellows.

Further information about the institute can be found on our website (www.ordwayresearch.org) Please forward inquiries including cover letter and resume to HR@ordwayresearch.org

Ordway Research Institute is an equal opportunity employer.

SP88912R

**Read Naturejobs
regularly.**

For better career
prospects.



naturejobs



Colgate University – Tenure Track Faculty Positions

Biology – Molecular Genetics

Department of Biology seeks a tenure-stream Assistant Professor to start August 2007. PhD or expectation of completion this academic year required; teaching and postdoctoral research experience desirable. The successful candidate will contribute to a foundation course called Molecules, Cells, and Genes, teach elective courses including behavioral genetics, and participate in University-wide programs. The appointee will join a biology faculty deeply committed to a strong, research-oriented program involving undergraduate students and will add to this effort by offering a research tutorial in their area of interest; opportunities also exist to lead a semester-long program at the NIH. Applications are especially encouraged from candidates whose research is focused in the area of behavioral genetics of model organisms. Please forward a letter of application with curriculum vitae, transcripts, and separate statements of teaching philosophy and research interests to Dr. Ken Belanger, Department of Biology, Colgate University, Hamilton, NY 13346-1398 and also arrange to have three letters of recommendation sent to this address. Review of applications will begin October 16 and continue until the position is filled. We intend to begin interviewing candidates by the beginning of November.

Biology – Plant or Wetlands Ecology

The Department of Biology seeks a tenure-stream Assistant Professor to start August 2007. PhD or expectation of completion this academic year required; teaching and postdoctoral research experience desirable. The successful candidate will: contribute to a foundation course in Evolution, Ecology, and Diversity; teach elective courses in botany/physiology and in their area of specialty; and contribute to interdisciplinary and University-wide programs (including Environmental Studies). The appointee will join a biology faculty deeply committed to a strong, research-oriented program involving undergraduate students and will add to this effort by offering a research tutorial in their area of interest. Please forward a letter of application with curriculum vitae, transcripts, and separate statements of teaching philosophy and research interests to Dr. Timothy McCay, Department of Biology, Colgate University, Hamilton, NY 13346-1398 and also arrange to have three letters of recommendation sent to this address. Review of applications will begin October 16 and continue until the position is filled. We intend to begin interviewing candidates by the beginning of November.

Psychology – Perception/Cognition

The Department of Psychology at Colgate University announces a tenure-track position in the area of (child/adult) Perception/Cognition at the rank of Assistant Professor beginning July, 2007. Completion of Ph.D. prior to or shortly after the date of hire is expected. Candidates with research interests in perceptual/cognitive development are of particular interest to us. The teaching load is equivalent to five courses per year which includes credit for supervising senior thesis research. Teaching responsibilities include Statistics, portions of staff-taught courses (e.g. Introductory Psychology, Research Methods), as well as upper-level courses in the candidate's specialty area which should include courses in Sensation and Perception, and possibly Developmental Psychology. Regular participation in all-University instructional programs is expected. Applicants should send a cover letter, curriculum vita, statements of teaching and research interests and three letters of recommendation, to Search Committee, Department of Psychology, Colgate University, Hamilton, NY 13346. Review of applications will begin on October 15, 2006.

Geology – Structural Geology/Tectonics

The Geology Department at Colgate University invites applications for an anticipated tenure-stream position at the rank of Assistant Professor in the field of structural geology/tectonics, beginning August 2007. We seek an individual with a Ph.D. who is committed to excellence in research and teaching at the undergraduate level. The area of specialization is open, but we are particularly interested in candidates whose research is field oriented. The successful applicant will teach Structural Geology, Tectonics, and develop other courses at the introductory level for non-majors and at the upper-level for geology students. Participation in the Geology Department's summer field course, involvement of undergraduates in research, and a willingness to contribute to other all-university curricula, such as the Scientific Perspectives program in the university's Core Curriculum, are expected.

Colgate University is a highly selective undergraduate liberal arts college with 2800 students. The Geology Department comprises nine faculty, a senior lecturer/lab instructor, and a technician. Analytical facilities include SEM-EDS, XRF, XRD, ICP-MS, GC-MS, AA, stable isotope mass spectrometer, micropaleontology lab, and geophysical equipment. The Geology Department maintains a web page at: <http://departments.colgate.edu/geology>

Review of applications will begin in September-early October, preliminary interviews will be held at GSA/Philadelphia, and on-campus interviews will begin shortly thereafter, pending status of the position at that time. Applications containing a curriculum vita, statements of teaching philosophy and research interests, and the names of three professional references should be sent to Dr. Constance M. Soja, Chair, Department of Geology, Colgate University, 13 Oak Drive, Hamilton, NY 13346-1398.

Colgate is a highly selective liberal arts college where both teaching and research are valued and supported. Colgate is located in the community of Hamilton in central Upstate New York. Potential applicants should visit our website at www.colgate.edu for further information about Colgate and the Hamilton area.

Colgate University is an equal opportunity/affirmative action employer. Developing and maintaining a diverse faculty and staff further the University's academic mission. Women and minorities are especially encouraged to apply.

SP88732R

UNIVERSITY OF ROCHESTER

The Department of Chemistry invites applications for positions in all areas of experimental chemistry at the Assistant, Associate, and Full Professor levels. Candidates with research interests in organic chemistry - broadly defined - are especially encouraged to apply. Candidates are expected to establish an outstanding program of original research and be effective teachers at the graduate and undergraduate levels. Applicants should send a curriculum vitae, a statement of research plans, teaching interests, and arrange for three letters of recommendation to be sent, preferably in electronic form, to Ms. Karen Dean, dean@chem.rochester.edu or mail to Chemistry Faculty Search Committee, c/o Ms. Karen Dean, Department of Chemistry, University of Rochester, RC Box 270216, Rochester, New York 14627-0216. Review of applications will begin on October 16, 2006.

The University of Rochester is an equal opportunity employer. Women and minority candidates are strongly encouraged to apply.

SP87917R



Aggressively recruiting the necessary scientific talent to reinforce its ranks, Roswell Park Cancer Institute is solidifying its position as a major cancer research powerhouse.

Alex Adjei, MD, PhD, has accepted the positions of Senior Vice President of Clinical Research and Chairman of the Department of Medicine, and Kelvin P. Lee, MD, accepted the positions of Chairman of the Department of Immunology and Vice Chairman of the Department of Medicine (Hematologic Malignancies). Both will assume their roles in October 2006.

Dr. Adjei, who comes from the Mayo College of Medicine, Rochester MN, has distinguished himself as a national leader in translational research, drug development and thoracic oncology. At RPCI, he will be responsible for expanding phase I and thoracic oncology programs; overseeing the Institute's planned Translational Oncology Core Laboratory; as well as assisting in the development of a new Clinical Research Unit. He is expected to add two MD/PhD and three PhD positions to his team.

Dr. Kelvin Lee, who comes from UM/Sylvester Cancer Center, University of Miami Miller School of Medicine, has been charged with the responsibility of developing strong scientific and clinical based collaborations to further RPCI's translational research agenda. He is expected to add two MD/PhD and five PhD positions to his research team.

In addition, RPCI is recruiting for a new chair of its Stress Biology program. Several new basic research faculty slots will be available within this program.

SP88987R

Bringing jobseekers and recruiters together.

Academic • Industry • Government



events • announcements • prospects •
regions • careers & recruitment
spotlight • special reports •
job channels • focus • movers

Search now at **naturejobs.com**

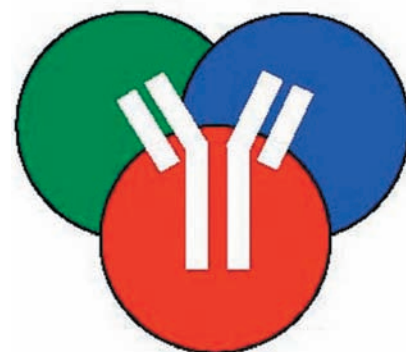
naturejobs
making science work

nature publishing group **npg**

9th Annual Upstate New York Immunology Conference

October 15-18, 2006

The Sagamore Resort, Bolton Landing, NY



SPOTLIGHT ON UPSTATE NEW YORK

Keynote Speakers:

David M. Mosser, Ph.D., University of Maryland

Marc K. Jenkins, Ph.D., University of Minnesota Medical School

The Upstate New York Immunology Conference (NYIC) has become recognized as a major regional immunology meeting, drawing participants from across upstate New York and the surrounding areas. The NYIC provides a forum for faculty from the various immunology programs to exchange information and ideas. The meeting also provides graduate students and research fellows a venue to present their research projects in a highly critical, yet supportive and nurturing environment. Session topics for the 2006 meeting include Immunity to Non-viral Infections, Tumor Immunity, Advances in Scientific Technology, Biodefense and Emerging Diseases, Mucosal Immunology, Immunity to Viral Infections, Antigen Recognition, and Cytokines and Immune Regulation. The annual NYIC is held at the Sagamore Resort on the shores of Lake George in the Adirondack Mountains, located approximately one hour from Albany, New York. The meeting is held during the prime period for viewing fall foliage.

Sponsoring Institutions: Albany Medical College, Cornell University, Roswell Park Cancer Institute, SUNY Upstate Medical University, University at Buffalo, University of Rochester Medical Center, Trudeau Institute, and New York State Department of Health Wadsworth Laboratories



Major financial supporters:

**National Institute of Allergy
and Infectious Diseases**

Taconic

BD Biosciences

Invitrogen

Information: www.amc.edu/NYIC/index.html

SP88911E

Dr Williamson and the master speed

Location, location, location.

Michael Livingston

*No speed of wind or water rushing by
But you have speed far greater. You can
climb
Back up a stream of radiance to the sky,
And back through history up the stream
of time.*

Robert Frost, *The Master Speed* (1934)

For immediate release:

INVESTIGATION INTO WILLIAMSON DISAPPEARANCE CONCLUDES
Princeton, N.J. — April 20, 2012 — Princeton Borough Police Chief Julian Bragg today announced that the investigation into the disappearance of Albert Williamson was closed, after nearly three weeks of intense and often secretive casework. Dr Williamson, a prominent professor of physics in the School of Natural Sciences at the world-renowned Institute for Advanced Study, who disappeared from his basement laboratory at approximately 2:30 in the afternoon of 1 April, has now been officially pronounced dead, a formal conclusion that many in the community had long accepted as inevitable.

"We can move on now, I suppose," said Williamson's colleague, Dr Aleksandar Restovic. "At last we can put Albert to rest — metaphorically, I mean, as recovery of the body obviously isn't feasible."

In a presentation complete with several charts and graphs, accompanied by an investigation team that included Restovic, two other members of Williamson's department at the institute, and Dr Beverly Leitner, the state's chief psychologist, Police Chief Bragg made clear that the root cause of the professor's death was the very thing that had propelled him on his meteoric rise through the physics profession: his single-mindedness.

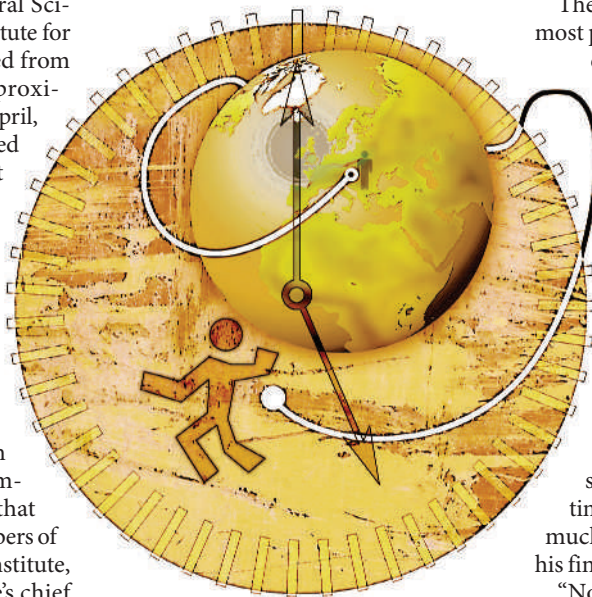
"The all-absorbing intensity with which Dr Williamson approached his work," Leitner went on to explain, "contributed not only to his extreme egocentricity, but also to his inability to look beyond the immediate problem at hand. For much of his career, these natural blinders to the surrounding issues of the world were a benefit, allowing him to break new ground in areas long since abandoned by the majority of mainstream scientists. Sadly, it also led to his death."

The investigators revealed that the

project that had been consuming so much of the physicist's life in the months before his death, and which he had held in such an unusual amount of secrecy that numerous strange and often fanciful rumours about it had begun to arise even before he vanished, was time-travel. The professor had quietly built, within his laboratory, a machine that could produce what Restovic termed a 'temporal manipulating field'.

"Through uncompromising determination," he explained, "Albert had managed to produce a device through which a person could journey to the same spot at a previous moment in time."

Investigators were able to determine that Williamson had made the sudden, and ultimately unfortunate, decision to test the machine himself on April 1, shortly after arguing with his colleagues that morning.



"When we got into work, Al kept bragging about having just completed a time machine," said Dr H. T. Zhang, another of Williamson's colleagues at the institute. "We figured it was his attempt at an April fools' prank, but he was insistent it was true and so we all argued about it. Things got pretty intense when Sanjay [Patel] said Al was 'not even wrong.'" Leitner believes that it was Williamson's intention to prove his colleagues wrong by using his device in order to join them, and himself, for lunch once more.

"At some level he recognized an inability to socialize with others," she said. "He figured the only person capable of communicating well with himself would be another

Albert Williamson. And although his relentless drive to accomplish this rather egocentric goal helped him to succeed in his seemingly quixotic efforts at time-travel, his hopes were doomed to fail."

The failure of Williamson's experiment, as the presentation made clear, was in his thinking so much about time and not enough about space.

"We remain fully convinced that Williamson's machine worked: the professor travelled back in time two-and-a-half hours, just as he'd planned," Bragg stated. "Yet in his insistence to solve the time-travel problem — a lack of clear thinking compounded by his emotional reaction to arguing with his peers — it simply never occurred to him that he ought to think about the undeniable relationships of time with space."

The Universe is not the static thing that most people assume it to be, the scientists explained. The speed of Earth's rotation at Princeton's latitude, for instance, is approximately 380 metres per second. Meanwhile, Earth is orbiting the Sun at 30 kilometres per second; the Sun is travelling through the Milky Way at 250 kilometres per second; and even the Galaxy itself is moving through space at approximately 300 kilometres per second. Thus, when Dr Williamson passed through his time machine, intending to leap backwards in his laboratory two-and-a-half hours, he successfully made it to the intended time, but not the intended place: Earth, much less his lab, had not yet arrived at his final destination.

"No doubt Albert felt much elation in the instant before his body was flash-frozen in the void of space," Restovic said.

In answer to a question regarding the two-and-a-half-hour delay between the decision to use the machine and the professor's disappearance into it, Bragg stated that the investigators concluded that Williamson wanted to be sure he would have an appetite for his second lunch. "He probably thought it would be rude to talk to himself while eating if he wasn't eating himself."

Michael Livingston has published widely in and out of academia, and he currently serves as an assistant professor of English at The Citadel in Charleston, South Carolina. His website can be found at <http://michael.d.livingston.googlepages.com>.

JACEY

Abstractions

FIRST AUTHOR

Today two-thirds of US citizens are overweight or obese. For the past 20 years, endocrinologist Michael Schwartz at the University of Washington School of Medicine has been unravelling the molecular and cellular mechanisms behind this trend. On page 289 of this issue, he reviews the brain's role in weight control. Schwartz talks to *Nature* about his work.

Are you an endocrinologist or a neuroscientist?

To be active or competitive in any scientific field, you have to wear a lot of different hats. Obesity research is a multidisciplinary endeavour. People come to it from different backgrounds, but none will be optimal on its own.

How did you become interested in obesity?

When I finished my medical training at Rush Medical College in Chicago in 1987, I decided to become an endocrinologist. I did a fellowship with Daniel Porte at the University of Washington, who was interested in the relationship between insulin signals in the brain and diabetes. He became convinced that body weight is regulated by hormones acting on the brain. I was captivated by the idea that obesity is a problem of regulation rather than will power.

How has the patient population changed since then?

We see heavier and more medically complicated patients. It is becoming routine to see patients over 227 kilograms in weight.

Has scientific progress kept up with the rise in obesity?

We have made tremendous progress in our scientific understanding of obesity. The delay is in translating this knowledge into clinical practice. It is clear that the integrity of the brain circuits that regulate body weight can be compromised by a nutrient-excess environment. If cells are continuously presented with an excess of glucose or other nutrients, this will cause a dysfunction in many cell types, undermining the sensitivity of brain circuits.

Are you optimistic that the obesity trends can be reversed?

What I envisage in the next ten years is that instead of having two slightly effective drugs there will be ten different types. Each drug may not be all that powerful on its own, but we will be able to use combinations of drugs to target compensatory pathways.

When you take a bite of food, do you think about what is happening in your brain?

Yes of course. After spending so long in this field there is a part of my brain that is always thinking about that. But for me it is more out of intellectual curiosity than concern. ■

MAKING THE PAPER

Zeresenay Alemseged

Perseverance yields the discovery of an infant hominin skeleton.

The Afar basin in northeastern Ethiopia has proved to be a fertile hunting ground for palaeontologists. In 1974, the region yielded the first remains of the genus *Australopithecus* in the shape of Lucy (*A. afarensis*). But the region is not the easiest place to conduct research, as it is caught up in conflicts between local tribes.

Despite the potential risks, Zeresenay Alemseged was determined to dig in the area — and his efforts have been richly rewarded (see page 296). Alemseged, a palaeoanthropologist at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, is a native Ethiopian, and he set about winning the support of the Afar locals as guides and guards to help him in his work.

The area Alemseged focused on is known as Dikika, and is some 10 kilometres from where Lucy was discovered. In late 2000, his team was lucky enough to spot a skull sticking out of the sand. It turned out to be a very significant find. It was the intact skull and a partial skeleton of a 3-year-old *A. afarensis* girl: the most complete, earliest specimen ever found, and it promises to supplant the 3.2-million-year-old Lucy for the title of the world's most famous hominin remains.

The sandstone preserved the face and a cast of its skull cavity — allowing researchers a glimpse of her appearance and a measure of brain volume. That first foray also revealed shoulder blades, collar bones, ribs and spinal column, hinting that there was more to unearth.

After depositing the first batch of fossils at the National Museum of Ethiopia, Alemseged returned to Dikika each year. Those subsequent trips yielded a humerus, fingers, both knee caps, the thigh and shinbone from both legs, and an almost complete foot, representing about 50% of the entire skeleton.



Besides annual rains that sweep aside sediment, Alemseged relied on other researchers and assistants to probe the sands. "You simply can't do this alone — you need a multidisciplinary team of scientists," he says.

Between trips, Alemseged separated and categorized the different elements of the skeleton. "For the past five years, I've spent most of my summer days describing the fossil and cleaning it under a microscope with dental instruments, because I decided not to use acid treatments that could destroy it," he says.

His preliminary analyses will surely spark fresh debates over hotly contested issues. For example, was the species a tree climber? The shoulder joint indicates that the arm could have been raised above the shoulder — necessary for climbing. And the discovery of the hyoid, the horseshoe-shaped bone that supports the tongue, is the first evidence for a voice box in *A. afarensis*.

As well as analysing the specimen, Alemseged plans to name the girl to honour all those involved. "Given the conflicts in the region, I'd like to build consensus among the locals for a name that represents peace," he says. ■

KEY COLLABORATIONS

How do you build a simple model of complicated processes? For Ping Chang, a climate scientist at Texas A&M University in College Station, it involved drawing on a diverse group of specialists.

On page 324, the collaborators discuss the link between the Pacific El Niño and the Atlantic Niño. Chang's interest in this relationship was sparked after conversations with scientists at the International Research Institute for Climate and Society

at Columbia University in New York. "From their experience working on climate-related issues in Africa, I learnt how important it is to understand and to predict Atlantic Niño," Chang says.

But unravelling how two geographically separate processes affect each other was not easy. Chang turned to graduate student Yue Fang, who built on the success of a climate model from the US National Center for

Atmospheric Research (NCAR) in Boulder, Colorado.

Working with Howard Seidel and Link Ji, Fang ironed out some of the potential glitches in the model. Ramalingam Saravanan helped the team to couple the NCAR atmospheric model to a simplified ocean model, allowing them to achieve their goal.

"Saravanan not only contributed intellectually to this work, but also provided assistance in statistical analysis of the data," says Chang. ■